Effective-mass reproducibility of the nearest-neighbor sp^3s^* models: Analytic results

Timothy B. Boykin
Department of Electrical and Computer Engineering, The University of Alabama in Huntsville, Huntsville, Alabama 35899

Gerhard Klimeck, R. Chris Bowen, and Roger Lake
Corporate Research and Development, Texas Instruments Incorporated, Dallas, Texas 75265
(Received 16 December 1996; revised manuscript received 7 March 1997)

We derive the exact expressions for the effective masses of the conduction and *all three* hole bands at Γ in the nearest-neighbor sp^3s^* model, both with and without the spin-orbit interaction. From these expressions we find that the nearest-neighbor sp^3s^* model can usually fit the electron mass better than the light- and heavy-hole masses and will *not* be equally successful across a broad spectrum of materials. [S0163-1829(97)03432-2]

It has been recognized for some time now that empirical tight-binding techniques are well-suited to modeling quantum heterostructures such as resonant tunneling diodes (RTD's) and quantum wells (QW's) since these methods can *potentially* reproduce important band-structure features better than either effective-mass or $\mathbf{k}\cdot\mathbf{p}$-type approaches can. The widespread recognition of this potential has unfortunately led many workers to mistakenly believe that the mere use of a multiband tight-binding model guarantees accurate results. Specifically, the band structures of the constituent materials in far too many applications of the tight-binding technique to heterostructures are *not* well reproduced, suggesting that there is little understanding of the band-structure features important for heterostructure calculations. A careful examination of these issues reveals that a *global* fit of the band structure is almost never desirable, for such a fit necessarily results in broad compromises. In contrast, good heterostructure calculations demand that certain energy gaps *and* effective masses be well modeled, even at the expense of the others, which are generally of little importance. The proper application of the tight-binding technique thus requires that we first determine *which* multi band tight-binding model will accurately reproduce the band-structure features of interest, and second, parametrize it in that way.

Neither of these issues has received much attention in the literature in spite of its importance. Most workers mploy the nearest-neighbor sp^3s^* model[1] because it is computationally convenient and can better fit the onduction band than can its sp^3 sibling, without having investigated its completeness. Thus, the capabilities nearest-neighbor sp^3s^* model are not well-determined, and it is generally believed to be sufficient in all cases of interest. Compounding this problem is the fact that both manual and automated fits tend to ignore the effective masses: even the very recent efforts at automatic parameter generation, such as those of Starrost *et al.*[2] make no effort to reproduce them.

A full understanding of the issues involved in choosing the best tight-binding model for a given problem demands that we investigate the band-structure features which are and are not well reproduced, and, more importantly, *why* this is so: such a study is particularly relevant for the nearest-neighbor sp^3s^* model due to its widespread use. In carrying out such a study we will generally find some features which are not well fit; this may occur for ne of two reasons. First, it might be that we simply did not find the very best parametrization since there are often many adjustable parameters (13 in the conventional implementation of the no-spin-orbit nearest-neighbor sp^3s^* model[1]). The second possibility is far more troublesome: namely, that there might be properties *intrinsic* to the model which make it difficult to fit a given feature or to simultaneously fit certain combinations of features. This is particularly true of the effective masses and the only way to settle this matter is to derive and study analytic expressions for them. Here we derive and study the effective-mass expressions for the nearest-neighbor sp^3s^* model, both with and without spin-orbit coupling. To our knowledge, neither have these formulas been previously published nor have their implications been discussed. From these formulas we find that there are indeed properties intrinsic to *both* models which can adversely affect our ability to simultaneously fit all of the gaps and masses necessary for many heterostructure problems.

We consider the nearest-neighbor sp^3s^* Hamiltonian both with and without the spin-orbit interaction; as is customary, we set the following parameters to zero: V_{s^*a,s^*c}, V_{sa,s^*c}, $V_{s^*a,sc}$.[1] The Hamiltonian operator is denoted H_0 for the model without spin-orbit coupling, and $H_0+H_{\rm SO}$, ($H_{\rm SO}$ represents the spin-orbit interaction) for the model with it. The resulting 10×10 matrix H_0 is reproduced in Ref. 1 so we do not repeat it here. We employ Chadi's[3] approach and notation when including the spin-orbit interaction. At the Γ point, H_0 is block diagonal, the subspaces being either one or two dimensional: $\{|s^*a\rangle\}$, $\{|s^*c\rangle\}$, $\{|sa\rangle,|sc\rangle\}$, $\{|xa\rangle,|xc\rangle\}$, $\{|ya\rangle,|yc\rangle\}$, $\{|za\rangle,|zc\rangle\}$. Note that the last three subspaces are degenerate. When the spin-orbit interaction is included, the 20×20 Hamiltonian first separates into 10×10 blocks, one in the basis $\{|s\mu;\uparrow\rangle, |s^*\mu;\uparrow\rangle, |z\mu;\uparrow\rangle, |x\mu;\downarrow\rangle, |y\mu;\uparrow\rangle\}$, $\mu=a,c$, the other with the spins reversed. Since $H_{\rm SO}$ couples only the p-like orbitals on the same site, we may construct new basis states by diagonalizing it in the subspace $\{|z\mu;\uparrow\rangle, |x\mu;\downarrow\rangle, |y\mu;\downarrow\rangle\}$ to obtain new eigenstates and energies ($\mu=a$ or c):[4]

TABLE I. Energies and coefficients for the bands in terms of the generic notation of Eqs. (4)–(7). The ''+'' solutions correspond to conduction bands, the ''−'' solutions to valence bands. The names ''electron'' and ''hole'' refer to both the ''+'' and ''−'' solutions of their respective 2×2 matrices. Thus, the lowest conduction band is ''$e+$,'' the light-hole valence band is ''$lh-$,'' etc. The notation ''both'' refers to both the spin-orbit and no-spin-orbit models while the exclusive designations ''s-o'' and ''no s-o'' refer to the spin-orbit and no-spin-orbit models, respectively. The subscript μ refers to anions (a) or cations (c).

Quantity in Eqs. (4)–(7)	Electrons (both)	Holes (no s-o)	Light holes (s-o)	Heavy holes (s-o)	Split off holes (s-o)
E_μ	$E_{s\mu}$	$E_{p\mu}$	$E_{p\mu}+\lambda_\mu$	$E_{p\mu}+\lambda_\mu$	$E_{p\mu}-2\lambda_\mu$
V	$V_{s,s}$	$V_{x,x}$	$V_{x,x}$	$V_{x,x}$	$V_{x,x}$
$E_\pm$	$E_\pm^{(e)}$	$E_\pm^{(h)}$	$E_\pm^{(\mathrm{lh})}$	$E_\pm^{(\mathrm{lh})}$	$E_\pm^{(\mathrm{soh})}$
ν_μ^+	σ_μ^+	ϱ_μ^+	$\varrho_\mu^{l,+}$	$\varrho_\mu^{l,+}$	$\varrho_\mu^{\mathrm{so},+}$
Δ	$\Delta^{(e)}$	$\Delta^{(h)}$	$\Delta^{(\mathrm{lh})}$	$\Delta^{(\mathrm{lh})}$	$\Delta^{(\mathrm{soh})}$

$$|h\mu;2\rangle=\frac{1}{\sqrt{2}}\,[|x\mu;\downarrow\rangle-i|y\mu;\downarrow\rangle],\quad \lambda_\mu, \tag{1}$$

$$|l\mu;2\rangle=\frac{1}{\sqrt{6}}\,[|x\mu;\downarrow\rangle+i|y\mu;\downarrow\rangle-2|z\mu;\uparrow\rangle],\quad \lambda_\mu, \tag{2}$$

$$|\mathrm{so}\mu;2\rangle=\frac{1}{\sqrt{3}}\,[|x\mu;\downarrow\rangle+i|y\mu;\downarrow\rangle+|z\mu;\uparrow\rangle],\quad -2\lambda_\mu, \tag{3}$$

(Similar expressions result for $\{|z\mu;\downarrow\rangle,\ |x\mu;\uparrow\rangle,\ |y\mu;\uparrow\rangle\}$, labeled ''1'' instead of ''2''.) Writing $H=H_0+H_{\mathrm{SO}}$, with the p orbitals replaced by Eqs. (1)–(3), we see that each 10×10 block again becomes block diagonal in one- and two-dimensional subspaces. For the states above, these are $\{|s^*a;\uparrow\rangle\}$, $\{|s^*c;\uparrow\rangle\}$, $\{|sa;\uparrow\rangle,\ |sc;\uparrow\rangle\}$, $\{|la;2\rangle,\ |lc;2\rangle\}$, $\{|ha;2\rangle,\ |hc;2\rangle\}$, $\{|\mathrm{so}a;\,2\rangle,\ |\mathrm{so}c;\rangle2\rangle\}$. When dealing with the spin-orbit model we shall from here on employ the $(\uparrow,2)$ states since identical results are obtained for the $(\downarrow,1)$ states.

In all cases, the 2×2 blocks have a common form, appearing in the subspace $\{|na\rangle,|nc\rangle\}$, where n is the orbital type, as

$$H_2=\begin{bmatrix}E_a & V\\ V & E_c\end{bmatrix},\quad V\in\mathbb{R}, \tag{4}$$

with eigenvalues,

$$E_\pm\equiv\overline{E}\pm\Delta,\quad \overline{E}\equiv\frac{E_a+E_c}{2},\quad E_\Delta\equiv\frac{E_a-E_c}{2},$$
$$\Delta\equiv\sqrt{E_\Delta^2+V^2} \tag{5}$$

and corresponding eigenvectors (+ and − refer to conduction and valence band, respectively),

$$|n\pm\rangle=\nu_a^\pm|na\rangle+\nu_c^\pm|nc\rangle, \tag{6}$$

$$\nu_a^\pm=\frac{\Delta+E_\Delta}{\sqrt{2}\sqrt{\Delta^2+\Delta E_\Delta}}=\nu_c^-,\quad \nu_c^+=\frac{V}{\sqrt{2}\sqrt{\Delta^2+\Delta E_\Delta}}=-\nu_a^-. \tag{7}$$

In Table I we list the parameters relevant to each 2×2 block and the notation for the eigenvalues and eigenvectors. Note that in the absence of the spin-orbit interaction all p-like bands have identical eigenvalues, Eq. (5), and coefficients, Eq. (7), whereas in the spin-orbit case the eigenvalues and coefficients of the split-off band differ from those of the light and heavy bands, which are identical.

Having obtained the eigenvalues and eigenvectors of H for both cases we now calculate the curvatures (inverse effective masses) at Γ. Different procedures are required for degenerate and nondegenerate bands.[5] Calculating the inverse masses (m_0/m_{zz}^*), where m_0 is the free-electron mass, and expressing the results in the notation of Table I, we find for the model without spin orbit

$$\frac{m_0}{m_e^*}=\frac{2m_0}{\hbar^2}\left(\frac{a}{4}\right)^2\left\{-\sigma_a^+\sigma_c^+V_{s,s}+\frac{[\sigma_a^+\varrho_a^+V_{sa,pc}-\sigma_c^+\varrho_c^+V_{pa,sc}]^2}{E_+^{(e)}-E_-^{(h)}}+\frac{[\sigma_a^+\varrho_c^+V_{sa,pc}+\sigma_c^+\varrho_a^+V_{pa,sc}]^2}{E_+^{(e)}-E_+^{(h)}}\right\}, \tag{8}$$

$$\frac{m_0}{m_{\mathrm{eh}}^*}=\frac{2m_0}{\hbar^2}\left(\frac{a}{4}\right)^2\left\{\varrho_a^+\varrho_c^+V_{x,x}+\frac{[\sigma_a^+\varrho_c^+V_{pa,sc}+\sigma_c^+\varrho_a^+V_{sa,pc}]^2}{E_-^{(h)}-E_-^{(e)}}+\frac{[\sigma_a^+\varrho_a^+V_{sa,pc}-\sigma_c^+\varrho_c^+V_{pa,sc}]^2}{E_-^{(h)}-E_+^{(e)}}\right.$$
$$\left.+\frac{(\varrho_a^+V_{s^*a,pc})^2}{E_-^{(h)}-E_{s^*a}}+\frac{(\varrho_c^+V_{pa,s^*c})^2}{E_-^{(h)}-E_{s^*c}}\right\}, \tag{9}$$

$$\frac{m_0}{m_{\mathrm{hh}}^*}=\frac{2m_0}{\hbar^2}\left(\frac{a}{4}\right)^2\left\{\varrho_a^+\varrho_c^+V_{x,x}+\frac{V_{x,y}^2}{E_-^{(h)}-E_+^{(h)}}\right\}, \tag{10}$$

where a is the conventional unit-cell cube edge and the subscripts e, lh, and hh refer to the electron, the light hole, and heavy hole, respectively. Note that the hole masses will be negative and that in the no spin-orbit model the heavy hole is doubly degenerate. For the spin-orbit model, the inverse masses are

$$\frac{m_0}{m_e^*}=\frac{2m_0}{\hbar^2}\left(\frac{a}{4}\right)^2\left\{-\sigma_a^+\sigma_c^+V_{s,s}+\left(\frac{2}{3}\right)\frac{[\sigma_a^+\varrho_a^{l,+}V_{sa,pc}-\sigma_c^+\varrho_c^{l,+}V_{pa,sc}]^2}{E_+^{(e)}-E_-^{(\mathrm{lh})}}+\left(\frac{1}{3}\right)\frac{[\sigma_a^+\varrho_a^{\mathrm{so},+}V_{sa,pc}-\sigma_c^+\varrho_c^{\mathrm{so},+}V_{pa,sc}]^2}{E_+^{(e)}-E_-^{(\mathrm{soh})}}\right.$$
$$\left.+\left(\frac{2}{3}\right)\frac{[\sigma_a^+\varrho_c^{l,+}V_{sa,pc}+\sigma_c^+\varrho_a^{l,+}V_{pa,sc}]^2}{E_+^{(e)}-E_+^{(\mathrm{lh})}}+\left(\frac{1}{3}\right)\frac{[\sigma_a^+\varrho_c^{\mathrm{so},+}V_{sa,pc}+\sigma_c^+\varrho_a^{\mathrm{so},+}V_{pa,sc}]^2}{E_+^{(e)}-E_+^{(\mathrm{soh})}}\right\}, \tag{11}$$

$$\frac{m_0}{m_{\mathrm{lh}}^*}=\frac{2m_0}{\hbar^2}\left(\frac{a}{4}\right)^2\left\{\varrho_a^{l,+}\varrho_c^{l,+}V_{x,x}+\left(\frac{2}{3}\right)\frac{[\sigma_a^+\varrho_c^{l,+}V_{pa,sc}+\sigma_c^+\varrho_a^{l,+}V_{sa,pc}]^2}{E_-^{(\mathrm{lh})}-E_-^{(e)}}+\left(\frac{2}{3}\right)\frac{[\sigma_a^+\varrho_a^{l,+}V_{sa,pc}-\sigma_c^+\varrho_c^{l,+}V_{pa,sc}]^2}{E_-^{(\mathrm{lh})}-E_+^{(e)}}\right.$$
$$\left.+\left(\frac{2}{3}\right)\frac{(\varrho_a^{l,+}V_{s*a,pc})^2}{E_-^{(\mathrm{lh})-}-E_{s*a}}+\left(\frac{2}{3}\right)\frac{(\varrho_c^{l,+}V_{pa,s*c})^2}{E_-^{(\mathrm{lh})}-E_{s*c}}+\left(\frac{1}{3}\right)\frac{V_{x,y}^2}{E_-^{(\mathrm{lh})}-E_+^{(\mathrm{lh})}}\right\}, \tag{12}$$

$$\frac{m_0}{m_{\mathrm{hh}}^*}=\frac{2m_0}{\hbar^2}\left(\frac{a}{4}\right)^2\left\{\varrho_a^{l,+}\varrho_c^{l,+}V_{x,x}+\left(\frac{1}{3}\right)\frac{V_{x,y}^2}{E_-^{(\mathrm{lh})}-E_+^{(\mathrm{lh})}}+\left(\frac{2}{3}\right)V_{x,y}^2\frac{[\varrho_a^{\mathrm{so},+}\varrho_a^{l,+}+\varrho_c^{\mathrm{so},+}\varrho_c^{l,+}]^2}{E_-^{(\mathrm{lh})}-E_+^{(\mathrm{soh})}}\right.$$
$$\left.+\left(\frac{2}{3}\right)V_{x,y}^2\frac{[\varrho_a^{\mathrm{so},+}\varrho_c^{l,+}-\varrho_c^{\mathrm{so},+}\varrho_a^{l,+}]^2}{E_-^{(\mathrm{lh})}-E_-^{(\mathrm{soh})}}\right\}, \tag{13}$$

$$\frac{m_0}{m_{\mathrm{soh}}^*}=\frac{2m_0}{\hbar^2}\left(\frac{a}{4}\right)^2\left\{\varrho_a^{\mathrm{so},+}\varrho_c^{\mathrm{so},+}V_{x,x}+\left(\frac{1}{3}\right)\frac{[\sigma_a^+\varrho_a^{\mathrm{so},+}V_{sa,pc}-\sigma_c^+\varrho_c^{\mathrm{so},+}V_{pa,sc}]^2}{E_-^{(\mathrm{soh})}-E_+^{(e)}}+\left(\frac{1}{3}\right)\frac{[\sigma_a^+\varrho_c^{\mathrm{so},+}V_{pa,sc}+\sigma_c^+\varrho_a^{\mathrm{so},+}V_{sa,pc}]^2}{E_-^{(\mathrm{soh})}-E_-^{(e)}}\right.$$
$$+\left(\frac{2}{3}\right)V_{x,y}^2\frac{[\varrho_a^{\mathrm{so},+}\varrho_c^{l,+}-\varrho_c^{\mathrm{so},+}\varrho_a^{l,+}]^2}{E_-^{(\mathrm{soh})}-E_-^{(\mathrm{lh})}}+\left(\frac{2}{3}\right)V_{x,y}^2\frac{[\varrho_a^{\mathrm{so},+}\varrho_a^{l,+}+\varrho_c^{\mathrm{so},+}\varrho_c^{l,+}]^2}{E_-^{(\mathrm{soh})}-E_+^{(\mathrm{lh})}}+\left(\frac{1}{3}\right)\frac{(\varrho_a^{\mathrm{so},+}V_{s*a,pc})^2}{E_-^{(\mathrm{soh})}-E_{s*a}}$$
$$\left.+\left(\frac{1}{3}\right)\frac{(\varrho_c^{\mathrm{so},+}V_{pa,s*c})^2}{E_-^{(\mathrm{soh})}-E_{s*c}}\right\}, \tag{14}$$

where in Eqs. (11)–(14) soh refers to the split-off hole. (The expressions for degenerate and nondegenerate bands have the same form due to our choice of basis, in which the degenerate perturbation matrices are already diagonal.) While these formulas will be useful to those fitting parameters for the two nearest-neighbor sp^3s^* models, the insight they provide into the capabilities and limitations of the models is of even greater importance. In particular, they explain certain trends which should be apparent to anyone fitting parameters for a wide range of materials.

Neither model seems to have an advantage at reproducing the electron mass, as is evident from the similarity of Eqs. (8) and (11). In the limit of small spin-orbit interaction, the second and third terms of Eq. (11) are together approximately equal to the second term of Eq. (8); this is also true of the fourth and fifth terms of Eq. (11) compared to the third term of Eq. (8). Not surprisingly, the second term of Eq. (8), the conduction-band–light-hole coupling, is usually the major contributor to the curvature: observe that since $V_{s,s}<0$ and $V_{xx}>0$, it follows that all of the ϱ_μ^+ (or $\varrho_\mu^{n,+}$) are positive, as is σ_a^+, but that σ_c^+ is negative. This implies that the numerator of the second term of Eq. (8) is usually significantly larger than that of the third term (we assume $V_{sa,pc}$, $V_{pa,sc}>0$ as usual). Taken together with the observation that the gap in the denominator of the second term of Eq. (8), the familiar band gap, is typically not too large and positive while that of the third term is generally larger and negative, we see that the second term ordinarily produces a large positive contribution whereas the third provides a small, negative one. [Similar reasoning applies to the second and third terms of Eq. (11) as compared to the fourth and fifth.] Because the conduction-band–light-hole coupling also appears in the light-hole mass expression, it will tend to limit our ability to independently fit the electron and light-hole masses along with the gap. Finally, note that due to the incompleteness of the basis the first term of each expression (the d^2H/dk^2 term) differs from what one would expect based on the standard $\mathbf{k}\cdot\mathbf{p}$ formula for the inverse mass.[5] This term (for both expressions) is always negative and, since $V_{s,s}$ is usually one of the largest-magnitude parameters, it can provide a significant negative contribution, increasing the electron mass.

The situation with respect to the light-hole masses in the two models, Eqs. (9) and (12), is another matter entirely, and here the spin-orbit model has a distinct advantage: its light-hole mass is usually larger (and sometimes much larger) than that of the no-spin-orbit model. To see why this is so, we compare Eqs. (9) and (12). As with the electron mass, the largest contribution (here negative) is the light-hole–conduction-band coupling. Negative contributions come as well from the light-hole-s^* couplings (these excited s-like

orbitals are high-lying); since $\varrho_c^+ > \varrho_a^+$ in most cases the s^*c term is often the second most significant of all. The positive contributions to the light-hole mass (first and second terms of each expression), unlike the negative contributions to the electron mass, do not generally reduce the curvature much. For the first (d^2H/dk^2) term this is because V_{xx} is usually one of the smaller parameters. For the second (light-hole to lowest valence-band coupling), this follows from the preceding discussion of the signs of the coefficients σ and ϱ and the generally large, positive, gap in the denominator. The key difference between the two models leading to a lower curvature in the spin-orbit case is that the second through fifth terms of Eq. (12) contribute at only two-thirds strength whereas they contribute fully in the no-spin-orbit model. While it is true that the last term of Eq. (12), the coupling to the ''heavy'' conduction band, is negative, and that $V_{x,y}$ can be sizeable, it is also the case that its denominator it typically significant. Moreover, it contributes at only one-third strength and most often does *not* compensate for the reduction in magnitude of the light-hole–conduction-band and light-hole-s^*c terms. Hence the heaver light-hole in the spin-orbit case.

The above analysis leads us to a startling conclusion: the model without spin-orbit coupling typically fails to reproduce accurately both the electron and light-hole effective masses. To see this, note that in both Eqs. (8) and (9) the electron-light-hole coupling typically accounts for most of the curvature. The difference in the two expressions is largely due to (i) the d^2H/dk^2 term, which significantly reduces the curvature of Eq. (8) but not of Eq. (9); and (ii) the light-hole-s^* couplings, which tend to increase the curvature of Eq. (9) but have no positive counterparts in Eq. (8). Thus, without the spin-orbit interaction, the light-hole-band curvature is typically *greater* than that of the conduction band and $|m_e^*| > |m_{\rm lh}^*|$, in contradistinction to what holds experimentally for most zinc-blende lattice materials.[6]

Including the spin-orbit interaction often ameliorates this situation, but it is *not* sufficient to guarantee that $|m_e^*| < |m_{\rm lh}^*|$: most intriguingly, reproducing this feature tends to be materials dependent. Observe that the electron couplings to both the light and split-off holes in Eq. (11) vary inversely with the gap at Γ (with respect to the latter term we assume small spin-orbit coupling), whereas only the electron-light-hole term in Eq. (12) does. When the gap becomes sufficiently large, the remaining negative terms of Eq. (12) can together become comparable in magnitude to the electron-split-off-hole term of Eq. (11); since the positive terms in the light-hole expression are generally smaller in size than the negative terms of the electron expression, it follows that here, too, we may find $|m_e^*| > |m_{\rm lh}^*|$. In practice, we have found this is the case for materials such as AlAs and AlSb, and without examining the expressions for the inverse masses it is quite unexpected. Rather surprisingly, then, we see that even when the spin-orbit coupling is included it will be difficult to reproduce $|m_e^*| < |m_{lh}^*|$ for materials with large gaps at Γ.

We may also encounter difficulties with the heavy-hole masses, Eqs. (10) and (13). In the limit of small spin-orbit interaction, the second and third terms of Eq. (13) together approximately equal the second term of Eq. (10). On first inspection it appears that the last term of Eq. (13), the coupling of the heavy- and split-off- hole bands, might be significant since its denominator is the often small spin-orbit splitting. The numerator, however, is even smaller, as we can see from some simple analysis. Substituting the values from Table I into this term we find

$$\frac{[\varrho_a^{\rm so,+}\varrho_c^{l,+} - \varrho_c^{\rm so,+}\varrho_a^{l,+}]^2}{E_-^{(\rm lh)} - E_-^{(\rm soh)}} = V_{x,x}^2 \frac{[\Delta^{(\rm soh)} - \Delta^{(\rm lh)} + \frac{3}{2}(\lambda_c - \lambda_a)]^2}{\Delta^{(\rm soh)} - \Delta^{(\rm lh)} + \frac{3}{2}(\lambda_c + \lambda_a)}. \quad (15)$$

Now, uniformly turning off the spin-orbit interaction by setting $\lambda_c = \alpha\lambda_a$ and taking the limit $\lambda_a \to 0$, we have

$$\frac{[\varrho_a^{\rm so,+}\varrho_c^{l,+} - \varrho_c^{\rm so,+}\varrho_a^{l,+}]^2}{E_-^{(\rm lh)} - E_-^{(\rm soh)}} \approx \frac{3}{2}\left[\frac{V_{x,x}^2(\alpha-1)^2(1+E_\Pi)^2}{(1-E_\Pi)+\alpha(1+E_\Pi)}\right]\lambda_a,$$

$$E_\Pi \equiv \frac{E_{pa} - E_{pc}}{\sqrt{(E_{pa}-E_{pc})^2 + 4V_{x,x}^2}}. \quad (16)$$

Thus, the last term of Eq. (13) is often quite small. What is perhaps most remarkable about the heavy-hole mass in either model is the relative *lack* of freedom we have in fitting it. Notice that in both Eqs. (10) and (13) the *only* nearest-neighbor parameters which affect the mass are V_{xx} (both directly and through the ϱ) and $V_{x,y}$. In contrast, the electron mass depends on $V_{x,x}$, $V_{s,s}$, $V_{sa,pc}$, and $V_{pa,sc}$, while the light-hole mass is a function of these plus $V_{s^*a,pc}$, and V_{pa,s^*c}, and, in the model with spin orbit, $V_{x,y}$ as well. While it is true that the freedom afforded by the s^* orbitals can be used to fit features at other points in the Brillouin zone, thus allowing us more leeway with the remaining orbitals, the above analysis makes it clear that we typically have little control over the heavy-hole mass in either model. These limitations, together with those associated with the light-hole mass, mean that we may encounter difficulties when using either model in valence- or inter-band heterostructure calculations.

The effective-mass formulas derived above and the conclusions drawn from them can aid us in heterostructure modeling; the most obvious use is of the formulas themselves in fitting the relevant effective masses. Our results on the hole masses reinforce the necessity of using the spin-orbit model for valence- or inter-band heterostructures. Interestingly, they have implications for conduction-band devices, too, the equal reproducibility of the electron mass notwithstanding. The light-hole mass especially cannot be altogether ignored since if it is incorrect, so too is the dispersion of the imaginary band linking the light hole and conduction bands at Γ. This imaginary band is important in determining the barrier attenuation of a RTD or QW, which in turn affects the resonances or energy levels; properly reproducing it requires good fits to the energy gap, electron-, and light-hole masses. Our remarks about the relative abilities of the spin-orbit and no-spin-orbit models to fit the both light-hole and electron masses suggest that we further investigate this subject.

As an example, let us consider AlAs, a common barrier material. In Table II we present the energy gap, light-hole,

TABLE II. Energy gap at Γ and effective masses of the conduction- and light-hole bands of AlAs reproduced by the nearest-neighbor sp^3s^* model both with and without the spin-orbit interaction. Experimental values are from Ref. 8.

Quantity	spin orbit	No spin orbit	Experiment
$\frac{m_e^*}{m_0}$	0.182	0.141	0.15
$\frac{m_{\rm lh}^*}{m_0}$	−0.135	−0.0786	−0.15
E_g (eV)	3.025	3.026	3.02

and electron masses reproduced by parametrizations of both the spin-orbit and no-spin-orbit sp^3s^* nearest-neighbor models, the parameters of which may be found in Table III. The complex bands of Fig. 1 were calculated using the generalized eigenproblem method[7] and the experimental values are from Ref. 8. As expected from the discussion about the masses for large-gap materials, we see that in both cases $|m_e^*|>|m_{lh}^*|$, the mismatch being much greater for the no-spin-orbit case. This mismatch in turn affects the imaginary band; the attenuation in the spin-orbit case is greater.

To see what effect this might have on a device we plot in Fig. 2 the transmission coefficient of a symmetric conduction-band GaAs/AlAs double-barrier heterostructure (16-ML barriers, 22-ML well) under flatband conditions. (The transmission resonances are helpful in evaluating optical devices such as multiple-quantum-well structures.) In order to isolate barrier attenuation effects we use the spin-orbit nearest-neighbor sp^3s^* model[1] for GaAs along with either the spin-orbit or no-spin-orbit nearest-neighbor sp^3s^* model[1] for AlAs and we restrict our attention to the first QW resonance. Although the calculations differ only in the AlAs model employed there is a significant disagreement in terms of background transmission and resonance width, and even some difference in resonance position. Thus, even for certain conduction-band heterostructures (particularly those incorporating materials with large gaps at Γ), we see that the spin-orbit model is probably the better choice.

Finally, we discuss an important limitation of the nearest-neighbor model used at the X points which makes it undesirable for AlAs transport (but not optical) device calculations; this limitation is most severe for the no-spin-orbit model. It has been previously remarked[9] that the bands are

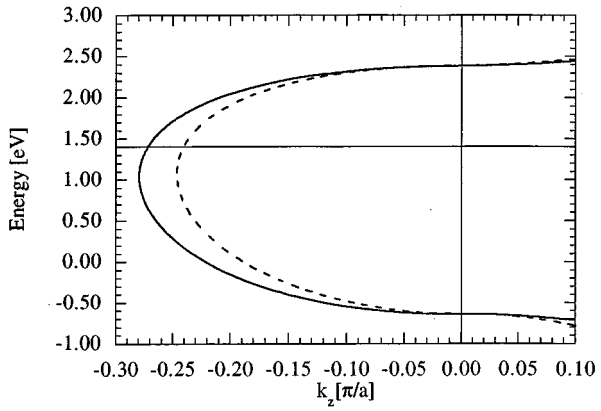

FIG. 1. Part of the complex band structure of AlAs ($T=300$ K) as reproduced by the nearest-neighbor sp^3s^* model with (solid line) and without (dashed line) the spin-orbit interaction. The real bands (conduction- and light-hole-valence-bands) are plotted in the $k_z>0$ part of the graph and the imaginary band linking them is plotted in the $k_z<0$ part of the graph. The solid horizontal line at around 1.4 eV indicates the approximate position of the GaAs conduction-band minimum; using it one may determine the attenuation seen by an electron tunneling through AlAs from the GaAs minimum.

flat in the transverse direction at the X points based on a symbolic evaluation of the determinant of $[H(\mathbf{k})-E]$. Here we give another demonstration: first, we evaluate the Hamiltonian matrix of Ref. 1 at $\mathbf{k}_X=(k_x,0,2\pi/a)$, that is, a vector ending on one of the diamond-shaped Brillouin-zone faces. Then, we make the following change of basis:

$$|y'a\rangle=-ic_x|ya\rangle+s_x|za\rangle,\quad |z'a\rangle=s_x|ya\rangle-ic_x|za\rangle, \tag{17}$$

$$|y'c\rangle=ic_x|yc\rangle+s_x|zc\rangle,\quad |z'c\rangle=s_x|yc\rangle+ic_x|zc\rangle, \tag{18}$$

where $c_x=\cos(k_xa/4)$, $s_x=\sin(k_xa/4)$. The Hamiltonian matrix in the new basis is now independent of k_x and so, therefore, are its eigenvalues, since they come from a polynomial equation independent of k_x. Hence, when the X-valley minimum occurs at the Brillouin-zone face the nearest-neighbor model (without spin orbit) will give an infinite transverse mass. The position of the X-valley minimum along [001] is, however, parameter dependent and so can occur before the zone boundary. Even in these cases, though, the transverse

TABLE III. Parameters for GaAs and AlAs in the nearest-neighbor sp^3s^* model; values are in eV.

Material	E_{sa}	E_{sc}	E_{pa}	E_{pc}	E_{s^*a}	E_{s^*c}	λ_a	λ_c
AlAs	−8.381 160	−1.744 670	0.229 440	2.832 840	6.730 574	5.972 840	0.140 00	0.008 00
AlAs[a]	−8.266 310	−1.782 020	0.344 290	2.947 690	6.845 424	6.087 690	0.000 00	0.000 00
GaAs	−8.510 704	−2.774 754	0.954 046	3.434 046	8.454 046	6.584 046	0.140 00	0.058 00

Material	$V_{s,s}$	$V_{sa,pc}$	$V_{pa,sc}$	$V_{s^*a,pc}$	V_{pa,s^*c}	$V_{x,x}$	$V_{x,y}$
AlAs	−6.664 20	5.600 00	6.800 00	4.220 00	7.300 00	1.878 00	3.860 00
AlAs[a]	−6.664 20	5.600 00	7.600 00	4.220 00	8.300 00	1.878 00	3.860 00
GaAs	−6.451 30	4.680 00	7.700 00	4.850 00	7.010 00	1.954 60	4.770 00

[a]The no-spin-orbit AlAs model.

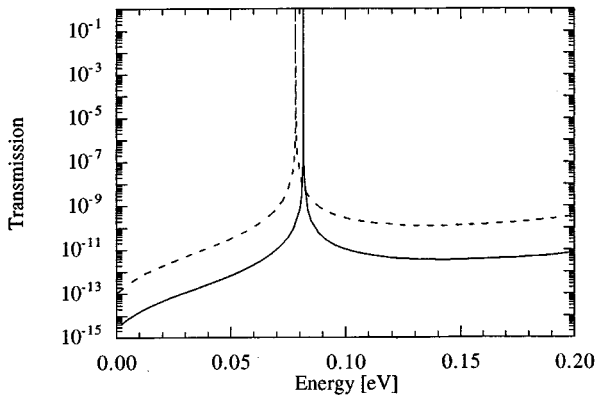

FIG. 2. Transmission-versus-energy graph under flatband (zero-bias, zero-charge) conditions for the GaAs/AlAs RTD discussed in the text. Both curves employ the spin-orbit model for GaAs. The solid curve uses the spin-orbit model for AlAs; the dashed curve uses the no-spin-orbit model for AlAs.

mass tends to be about an order of magnitude (or more) too large. (It is difficult to say much about the details of this parameter dependence since the subspace in question is six dimensional in the sp^3s^* model and even four dimensional in the sp^3 model.) In the model with spin-orbit interaction, the k-independent coupling of the p-like orbitals likewise inhibits a detailed discussion of its properties at the zone faces. Nevertheless, we have found that this model, too, tends to do a poor job of reproducing the X-valley transverse mass.

In conclusion, we have discussed the band-structure features of importance for transport and electronic structure calculations. Keeping these considerations in mind, we have examined several properties of the widely employed nearest-neighbor sp^3s^* model relevant to its use in describing semiconductor heterostructures. We have derived and presented exact expressions for the (inverse) effective masses of the conduction, light-hole, heavy-hole, and, in the spin-orbit model, split-off-hole bands at Γ. From these expressions we have drawn several interesting conclusions about the relative abilities of the spin-orbit and no-spin-orbit versions to fit these masses. We find that the no-spin-orbit model tends to reproduce $|m_e^*|>|m_{lh}^*|$, and that problems in fitting the light-hole mass tend to be worse with larger-gap materials. We have furthermore seen how this undesirable result can affect the attenuation of the barriers of a RTD or multiple-quantum-well structure, and thus transmission behavior and energies. We also find that there is relatively little freedom available for fitting the heavy-hole mass; again, this tendency does not seem to have been previously noticed. From these results, we can see that for some materials *neither* nearest-neighbor sp^3s^* model is really appropriate and more complete models are likely required. The inverse mass formulas we have derived are immediately useful to those fitting the parameters of the two sp^3s^* nearest-neighbor models either manually or by computer program. More importantly, these formulas can serve as guides to selecting which, if any, of these models is the more appropriate for a given purpose.

Note added. The paper of Loehr and Talwar[10] recently appeared in which are given the expressions for only the electron and heavy-hole masses in the no spin-orbit, second-nearest-neighbor sp^3 model. The expressions for *all* hole bands as well as the conduction-band in even the spin-orbit, second-nearest-neighbor sp^3s^* are readily obtainable using the methods employed here, contrary to the conjecture of Loehr and Talwar. We have already derived these formulas and will present them in a future study.

T.B.B. gratefully acknowledges the support of Texas Instruments, Inc.

[1] P. Vogl, Harold P. Hjalmarson, and John D. Dow, J. Phys. Chem. Solids **44**, 365 (1983). There are some minor typographical errors in the Hamiltonian matrix here: the geometrical factors g_1, g_2, and g_3 appearing in the $|s^*a\rangle$ column should be replaced by their respective complex conjugates.

[2] F. Starrost, S. Bornholdt, C. Solterbech, and W. Schattke, Phys. Rev. B **53**, 12 549 (1996).

[3] D. J. Chadi, Phys. Rev. B **16**, 790 (1977).

[4] See, for example, Supryio Datta, *Quantum Phenomena* (Addison-Wesley, Menlo Park, CA, 1989), pp. 210–13.

[5] Timothy B. Boykin, Phys. Rev. B **52**, 16 317 (1995).

[6] We remark that one might attempt to correct this problem by allowing the s^* orbitals to interact, i.e., taking $V_{s^*a,s^*c}\neq 0$. In order to reasonably reproduce the known bands at Γ while also increasing the light-hole mass under such a scheme, we would have to move the lower of the two s^* bands *below* the valence bands; given the high-lying free-atom s^* energies this would also greatly increase the energy of the upper s^* band. While this would indeed accomplish the purpose (by replacing one sizeable negative term with a smaller positive one and reducing the magnitude of the remaining negative term), it would also tend to significantly distort the bands elsewhere (e.g., along [001]) due to the size of the V_{s^*a,s^*c} parameter required. Remembering that the original intent of the s^* orbitals was to avoid d orbitals, we see that such a plan ought to be rejected as physically unreasonable.

[7] Timothy B. Boykin, Phys. Rev. B **54**, 8107 (1996).

[8] Sadao Adachi, J. Appl. Phys. **58**, R1 (1985).

[9] Timothy B. Boykin, Jan P. A. van der Wagt, and James S. Harris, Jr., Phys. Rev. B **43**, 4777 (1991).

[10] J. P. Loehr and D. N. Talwar, Phys. Rev. B **55**, 4353 (1997).