Valence band effective-mass expressions in the sp^3d^5s* empirical tight-binding model applied to a Si and Ge parametrization

Timothy B. Boykin

Department of Electrical and Computer Engineering, The University of Alabama in Huntsville, Huntsville, Alabama 35899, USA

Gerhard Klimeck* and Fabiano Oyafuso

Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Road, MS 169-315, Pasadena, California 91109, USA

Exact, analytic expressions for the valence band effective masses in the spin-orbit, sp^3d^5s* empirical tight-binding model are derived. These expressions together with an automated fitting algorithm are used to produce improved parameter sets for Si and Ge at room temperature. Detailed examinations of the analytic effective-mass expressions reveal critical capabilities and limitations of this model in reproducing simultaneously certain gaps and effective masses. The [110] masses are shown to be completely determined by the [100] and [111] masses despite the introduction of d orbitals into the basis.

I. INTRODUCTION

Heterostructures with atomically sharp interfaces lie at the heart of advanced semiconductor devices employed for electronic computation and optical detection and emission. While such heterostructures have typically been explored in the III-V and II-VI material systems, commercial device scaling in the Si, Ge, and SiGe material systems has also reached critical device dimensions of a few tens of nanometers. Of particular interest is also the development of Si-based quantum dot heterostructures for quantum computing.[1,2]

Proof-of-principle device modeling has typically been performed in continuum basis, plane wave approaches such as effective mass or $\mathbf{k} \cdot \mathbf{p}$. However, realistic modeling of technically relevant heterostructures requires the atomistic resolution of interface details, atomic disorder, the incorporation of finite device boundaries, large local electric fields, and coherent coupling between bands. Localized orbital approaches are attractive since they can model these effects without the introduction of an artificial periodicity or an extraordinarily large plane wave basis.

Any realistic electronic heterostructure model, whatever the underlying basis, must accurately reproduce the experimentally verified energy gaps and effective masses for the relevant bands. An approach which delivers a band gap that is off by a few tens of meV or effective masses that are off by tens of percent can properly model neither the energies of quantized states nor the confinement of these states. The strength of empirical tight-binding models is their ability to properly model such critical material properties with a localized, atomistic orbital basis. Empirical tight-binding techniques have therefore developed into the primary choice for many researchers interested in the quantitative modeling of electronic structure on a nanometer scale.

The past several years have seen a broader recognition of the requirements of proper band structure reproduction, and the result has been a series of improvements to the pioneering work of Slater and Koster.[3] A major enhancement was the addition of the excited s-like orbital, $s*$, by Vogl, Hjal-marson, and Dow,[4] which yielded better reproduction of the conduction band of diamond and zinc blende semiconductors along [100]. Much more recently, Jancu *et al.*[5] have further improved the method by adding excited d orbitals to a nearest-neighbor approach, resulting in superior reproduction of the X-valley transverse masses in these semiconductors.

The progress in the application of the tight-binding method to quantitative device modeling has been slowed by the one major handicap of the method: the difficult parameter fitting process. One of the strengths of continuum models such as effective mass and $\mathbf{k} \cdot \mathbf{p}$ is that they are directly parametrized by observables such as effective masses, band gaps, and band offsets. In contrast, empirical tight-binding and empirical pseudopotential methods are based on atomic interaction representations and need to be fitted to these bulk material properties. The fitting process is difficult on two levels: (i) the determination whether or not a particular tight-binding model can indeed properly reproduce some critical material properties, and (ii) the mapping of the large number of orbital coupling parameters onto the set of observables. The development of analytic expressions for effective masses for various tight-binding models,[6–10] along with automated fitting procedures,[11] has addressed some of these difficulties in tight binding.

Of all the observables, the effective masses invariably prove to be the most troublesome features to reproduce in either automated or manual fits. There are two reasons for this. First, the inverse effective mass is (up to constants) the second derivative of the band energy $E_n(\mathbf{k})$. Second, the masses are much more complicated functions of the tight-binding parameters than are most of the gaps. In those cases where they are available, exact, analytic formulas for the effective masses have not only aided in better parametrizations for tight-binding models, but have also provided much insight as to the suitability of different models for different materials. Results for the masses at Γ in the nearest-neighbor sp^3s* models[6] (with and without the spin-orbit coupling), the second-near-neighbor, spin-orbit sp^3s* model,[7,8] and the second- and third-near-neighbor no-spin-orbit[9,10] sp^3 models

have recently been published. Reference 8 in fact gives the valence band effective masses along the [110] and [111] directions, which, together with the [001] results from Refs. 6 and 7, demonstrate the capability of the sp^3s^* model to reproduce all of the valence band masses in many materials of interest.

Although both the nearest-neighbor, spin-orbit $sp^3d^5s^*$ model[5] and the second-near-neighbor, spin-orbit sp^3s^* model can accurately reproduce the observables of interest in heterostructures (effective masses at gaps at Γ, X, and L), there are important reasons to prefer the nearest-neighbor, spin-orbit $sp^3d^5s^*$ model. These reasons become apparent when one applies each model to strained nanostructures, for distortions are handled much more easily in a nearest-neighbor model. This theoretical difficulty has led the nano-electronic modeling tool (NEMO) three-dimensional (3D) development[12] to be limited to nearest-neighbor interactions, while NEMO 1D still handles nearest- and second-nearest-neighbor models.[13] To fully qualify the capabilities of the $sp^3d^5s^*$ model and to aid its parametrization, exact, analytic effective-mass formulas are highly desirable. These formulas for the valence band effective masses in the directions of [001], [110], and [111] are derived here. We have not been able to find analytic formulas for the conduction band masses at L and in the X valleys.

Section II shows how a series of basis transformations block-diagonalizes the Hamiltonian at Γ into blocks no larger than 2×2, which are sufficiently small that useful formulas for effective masses can be derived. The method of Ref. 15 is then utilized to derive these exact, analytic, inverse effective-mass formulas for the valance bands along [001], [110], and [111]. Section III presents the results of the automated $sp^3d^5s^*$ parameter generation for Si and Ge. Section IV connects the analytic and numerical work in the previous two sections and discusses the ability of the $sp^3d^5s^*$ model to reproduce faithfully and simultaneously all of these masses. Section V contains our conclusions.

II. METHOD

A. Block diagonalization of the Hamiltonian

The first step in the process is to diagonalize the spin-orbit, nearest-neighbor, $sp^3d^5s^*$ Hamiltonian[3] for diamond at Γ. The spin-orbit interaction is assumed to be a p-orbital, same-atom only parameter, as in Refs. 5 and 16. Following Ref. 5, a single onsite parameter for all d orbitals is employed, and only two-center integrals determine the nearest-neighbor parameters; the notation follows Ref. 3. Such restrictions are almost universal in the literature and are restated here for completeness. In the $sp^3d^5s^*$ basis, the $\mathbf{k}$ dependence of the Hamiltonian $\underline{\mathbf{H}}(\mathbf{k})$ is contained in the functions $g_i(\mathbf{k})$ (see Ref. 4 for definitions): $g_0(\mathbf{0})=1$ while $g_i(\mathbf{0})=0$, $i=1,2,3$. As a result, the Hamiltonian $\underline{\mathbf{H}}(\mathbf{0})$ as written is block diagonal, consisting of two 12-dimensional subspaces, two four-dimensional subspaces, and four two-dimensional subspaces. Since Si and Ge possess diamond symmetry, further reduction of the 12×12 blocks to 4×4 is possible; the basis transformations which accomplish this are discussed below.

These subspaces of the Hamiltonian as written are conveniently described in terms of zone-center Bloch sums of atomiclike orbitals. Since there are two atoms per unit cell in diamond there are two orbitals of a given type per cell, one on each atom. The 12-dimensional subspaces include both p- and d-like orbitals, one subspace consisting of orbital types $\{p_x \uparrow, p_y \uparrow, p_z \downarrow, d_{xy} \downarrow, d_{zx} \uparrow, d_{yz} \uparrow\}$, the other like it but with the spins reversed. The four-dimensional subspaces are purely s-like, one having orbitals $\{s \uparrow, s^* \uparrow\}$, the other with the spins reversed. Finally, there is one two-dimensional subspace for each of the orbital types $d_{x^2-y^2} \uparrow$, $d_{x^2-y^2} \downarrow$, $d_{3z^2-r^2} \uparrow$, and $d_{3z^2-r^2} \downarrow$.

The 12×12 blocks are reduced to 4×4 using the heavy-, light-, and split-off hole states which diagonalize the (spin-orbit) sp^3s^* tight-binding or the sp^3 $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian (see, for example, Ref. 6), here applied to the d-like states (substituting d_{yz} for p_x, etc.) also, as suggested by symmetry considerations. This choice is motivated by the fact that the spin-orbit Hamiltonian has the same matrix elements as in the sp^3s^* tight-binding model. These transformations are

$$|\omega h \mu; 1\rangle = \frac{1}{\sqrt{2}}[|\alpha \mu; \uparrow\rangle + i|\beta \mu; \uparrow\rangle], \quad (1)$$

$$|\omega l \mu; 1\rangle = \frac{1}{\sqrt{6}}[|\alpha \mu; \uparrow\rangle - i|\beta \mu; \uparrow\rangle + 2|\gamma \mu; \downarrow\rangle], \quad (2)$$

$$|\omega so \mu; 1\rangle = \frac{1}{\sqrt{3}}[|\alpha \mu; \uparrow\rangle - i|\beta \mu; \uparrow\rangle - |\gamma \mu; \downarrow\rangle], \quad (3)$$

$$|\omega h \mu; 2\rangle = \frac{1}{\sqrt{2}}[|\alpha \mu; \downarrow\rangle - i|\beta \mu; \downarrow\rangle], \quad (4)$$

$$|\omega l \mu; 2\rangle = \frac{1}{\sqrt{6}}[|\alpha \mu; \downarrow\rangle + i|\beta \mu; \downarrow\rangle - 2|\gamma \mu; \uparrow\rangle], \quad (5)$$

$$|\omega so \mu; 2\rangle = \frac{1}{\sqrt{3}}[|\alpha \mu; \downarrow\rangle + i|\beta \mu; \downarrow\rangle + |\gamma \mu; \uparrow\rangle], \quad (6)$$

where in Eqs. (1)–(6) orbital designators have been chosen as $(\omega; \alpha, \beta, \gamma) = (p; x, y, z)$, or $(\omega; \alpha, \beta, \gamma) = (d; yz, zx, xy)$, with spin labels $\{\uparrow, \downarrow\}$, and fcc sublattice label $\mu = \{a, c\}$. Note that the states with $\omega = p(d)$ are the $p(d)$-like components of the heavy- (h), light- (l), or split-off (so) subspaces to be extracted below.

The roots of the quartic equation are sufficiently complicated that to obtain useful analytic formulas one must further decompose all of the four-dimensional subspaces. In diamond the two atoms of a unit cell are identical, and further simplification results from changing to a basis of sums and differences of states of the same type on the two atoms, reminiscent of the bonding and antibonding orbitals of molecular physics. For the six subspaces resulting from the transformations in Eqs. (1)–(6), the following change of basis is performed:

TABLE I. Nearest-neighbor (NN) parameters appearing in the inverse mass formulas in terms of the Slater-Koster (Ref. 3) two-center integrals. Specific values of the two-center integrals for Si and Ge are given in Table IV below.

NN parameter	Slater-Koster	NN parameter	Slater-Koster
$V_{s,s}$	$4V_{ss\sigma}$	$V_{x,y}$	$\frac{4}{3}[V_{pp\sigma}-V_{pp\pi}]$
$V_{s,s*}$	$4V_{ss*\sigma}$	$V_{p,d}$	$\frac{4}{3}\left[V_{pd\sigma}-\frac{2}{\sqrt{3}}V_{pd\pi}\right]$
$V_{s*,s*}$	$4V_{s*s*\sigma}$	$U_{p,d}$	$\frac{4}{3}\left[V_{pd\sigma}+\frac{1}{\sqrt{3}}V_{pd\pi}\right]$
$V_{s,p}$	$\frac{4}{\sqrt{3}}V_{sp\sigma}$	$W_{p,d}$	$\frac{8}{3}V_{pd\pi}$
$V_{s*,p}$	$\frac{4}{\sqrt{3}}V_{s*p\sigma}$	$V_{d,d}$	$\frac{4}{9}[3V_{dd\sigma}+2V_{dd\pi}+4V_{dd\delta}]$
$V_{s,d}$	$\frac{4}{\sqrt{3}}V_{sd\sigma}$	$\tilde{V}_{d,d}$	$\frac{4}{3}[2V_{dd\pi}+V_{dd\delta}]$
$V_{s*,d}$	$\frac{4}{\sqrt{3}}V_{s*d\sigma}$	$U_{d,d}$	$\frac{4}{9}[3V_{dd\sigma}-V_{dd\pi}-2V_{dd\delta}]$
$V_{x,x}$	$\frac{4}{3}[V_{pp\sigma}+2V_{pp\pi}]$	$W_{d,d}$	$\frac{8}{3\sqrt{3}}[V_{dd\delta}-V_{dd\pi}]$

$$|p\nu n;\pm\rangle=\frac{1}{\sqrt{2}}[\pm|p\nu a;n\rangle+|p\nu c;n\rangle], \qquad (7)$$

$$|d\nu n;\pm\rangle=\frac{1}{\sqrt{2}}[|d\nu a;n\rangle\mp|d\nu c;n\rangle], \qquad (8)$$

where $n=\{1,2\}$, $\nu=\{h,l,so\}$. The remaining four-dimensional subspaces, spanned by the s and $s*$ orbitals (one subspace for each spin direction) are block diagonalized by the same procedure:

$$|\sigma\chi;\pm\rangle=\frac{1}{\sqrt{2}}[|\sigma a;\chi\rangle\mp|\sigma c;\chi\rangle], \qquad (9)$$

where $\sigma=\{s,s*\}$, $\chi=\{\uparrow,\downarrow\}$. The development that follows is simplified by a short-hand notation for two 20-dimensional subspaces, labeling the 12 states of Eqs. (1)–(3) above, in addition to the eight states (for both $\mu=\{a,c\}$) $\{|s\downarrow;\pm\rangle,|s*\downarrow;\pm\rangle,|(x^2-y^2)\mu;\downarrow\rangle,|(3z^2-r^2)\mu;\downarrow\rangle\}$ the "1 states," and the 12 states of Eqs. (4)–(6) above, together with the eight states $\{|s\uparrow;\pm\rangle,|s*\uparrow;\pm\rangle,|(x^2-y^2)\mu;\uparrow\rangle,|(3z^2-r^2)\mu;\uparrow\rangle\}$ the "2 states."

The foregoing basis transformations thus block-diagonalize the original 40×40 Hamiltonian into 2×2 matrices. In what follows below one obtains the same results for the 1 and 2 states, so in the interest of brevity only results for the 2 states are listed. The Hamiltonian blocks now take the same form (in a generic $\{|a\rangle,|b\rangle\}$ basis):

$$H_2=\begin{bmatrix}E_a & V\\ V & E_b\end{bmatrix} \qquad (10)$$

with eigenvalues

$$E_\pm=\bar{E}\pm\Delta, \qquad (11)$$

$$\bar{E}=\frac{1}{2}(E_a+E_b),\quad E_\Delta=\frac{1}{2}(E_a-E_b),\quad \Delta=\sqrt{E_\Delta^2+V^2} \qquad (12)$$

and corresponding eigenvectors

$$|+\rangle=\eta_a|a\rangle+\eta_b|b\rangle, \qquad (13)$$

$$|-\rangle=-\eta_b|a\rangle+\eta_a|b\rangle, \qquad (14)$$

$$\eta_a=\frac{E_\Delta+\Delta}{\sqrt{2}\sqrt{\Delta^2+\Delta E_\Delta}},\quad \eta_b=\frac{V}{\sqrt{2}\sqrt{\Delta^2+\Delta E_\Delta}}. \qquad (15)$$

The states themselves are found by substituting the quantities in Tables I and II into the above equations, except for the case of the two subspaces $\{|(x^2-y^2)a;\uparrow\rangle,|(x^2-y^2)c;\uparrow\rangle\}$, $\{|(3z^2-r^2)a;\uparrow\rangle,|(3z^2-r^2)c;\uparrow\rangle\}$, for which the results are so simple (due to the degeneracy of the atomiclike orbitals) that they are stated separately. Both subspaces have the same 2×2 block Hamiltonian and the resulting eigenvalues and eigenvector coefficients are, in the parlance of Eqs. (11) and (15),

TABLE II. Abbreviations used with Eqs. (10)–(15) to get specific eigenvalues and eigenvectors of the 2×2 blocks in terms of the tight-binding parameters. Slater-Koster (Ref. 3) expressions for on-site parameters and the abbreviations of Table I for nearest-neighbor parameters are used.

Abbreviation	Slater-Koster/Table I
$E_\pm^{(s)}$	$E_s \mp V_{s,s}$
$E_\pm^{(s^*)}$	$E_{s*} \mp V_{s*,s*}$
$E_\pm^{(l)}$	$E_p + \lambda \pm V_{x,x}$
$E_\pm^{(so)}$	$E_p - 2\lambda \pm V_{x,x}$
$E_\pm^{(d)}$	$E_d \mp V_{d,d}$

$$E_{\delta,\pm} = E_d \pm \tilde{V}_{d,d}, \quad \tilde{V}_{d,d} = \left(\frac{4}{3}\right)[2V_{dd\pi} + V_{dd\delta}], \quad (16)$$

$$\eta_{\delta,a} = \eta_{\delta,b} = \frac{1}{\sqrt{2}}. \quad (17)$$

These transformations diagonalize the Hamiltonian at zone center, providing the eigenstates with which the inverse effective masses can be calculated.

B. Inverse effective masses along [001]

The inverse effective masses for the valence bands are calculated using the method of Ref. 15. The heavy, light, and split-off holes are the four states $(n=1,2)$ $|hn;-;-\rangle$ (heavy), $|ln;-;-\rangle$ (light), and $|son;-;-\rangle$ (split off); see Table III for specific decompositions for the 2 states. Along [001] neither of the matrices $d\underline{\mathbf{H}}/dk_z$, $d^2\underline{\mathbf{H}}/dk_z^2$ couples the 1 states to the 2 states. Nor for that matter does either matrix couple the degenerate heavy- and light-hole states within each of the 20-dimensional (i.e., "1" or "2") subspaces. The inverse effective-mass formulas therefore take the same form as in the nondegenerate perturbation theory.

The contributions to the inverse mass are of two types:[15] the incompleteness correction (the second-order matrix $d^2\underline{\mathbf{H}}/dk^2$ to first order in perturbation theory) and couplings to other bands (the first-order matrix $d\underline{\mathbf{H}}/dk$ to second order in perturbation theory). It is useful to first list these contributions separately since many of them will appear in the [111] inverse mass formulas as well. The heavy and light holes have identical incompleteness corrections,

$$w_{(inc),l}^{[001]} = (\gamma_{d,l}^{(-)})^2 V_{x,x} - (\gamma_{p,l}^{(-)})^2 V_{d,d} - 2\gamma_{d,l}^{(-)}\gamma_{p,l}^{(-)}V_{p,d}, \quad (18)$$

while the split-off hole incompleteness term is

$$w_{(inc),so}^{[001]} = (\gamma_{d,so}^{(-)})^2 V_{x,x} - (\gamma_{p,so}^{(-)})^2 V_{d,d} - 2\gamma_{d,so}^{(-)}\gamma_{p,so}^{(-)}V_{p,d}. \quad (19)$$

Next come the couplings of the valence bands to other bands. The heavy valence band couples to both of the light conduction bands; with the same strengths the light valence band couples to the heavy conduction bands:

$$w_{v,l++} = \frac{2}{E_{l,-}^{(-)} - E_{l,+}^{(+)}}\left(\frac{1}{3}\right)[-\gamma_{d,l}^{(-)}\gamma_{p,l}^{(+)}V_{x,y} + \gamma_{p,l}^{(-)}\gamma_{d,l}^{(+)}U_{d,d}$$
$$+ (\gamma_{d,l}^{(-)}\gamma_{d,l}^{(+)} + \gamma_{p,l}^{(-)}\gamma_{p,l}^{(+)})U_{p,d}]^2, \quad (20)$$

$$w_{v,l+-} = \frac{2}{E_{l,-}^{(-)} - E_{l,-}^{(+)}}\left(\frac{1}{3}\right)[\gamma_{d,l}^{(-)}\gamma_{d,l}^{(+)}V_{x,y} + \gamma_{p,l}^{(-)}\gamma_{p,l}^{(+)}U_{d,d}$$
$$+ (\gamma_{d,l}^{(-)}\gamma_{p,l}^{(+)} - \gamma_{p,l}^{(-)}\gamma_{d,l}^{(+)})U_{p,d}]^2. \quad (21)$$

The other heavy valence band couplings are to the split-off conduction bands,

$$w_{v,so++} = \frac{2}{E_{l,-}^{(-)} - E_{so,+}^{(+)}}\left(\frac{2}{3}\right)[-\gamma_{d,l}^{(-)}\gamma_{p,so}^{(+)}V_{x,y}$$
$$+ \gamma_{p,l}^{(-)}\gamma_{d,so}^{(+)}U_{d,d} + (\gamma_{d,l}^{(-)}\gamma_{d,so}^{(+)} + \gamma_{p,l}^{(-)}\gamma_{p,so}^{(+)})U_{p,d}]^2, \quad (22)$$

TABLE III. New states, basis states, eigenvalues, and eigenvector coefficients for the eight two-dimensional subspaces reduced from the original 12- and four-dimensional subspaces in the notation of Tables I and II and Eqs. (1)–(15). The names of the $|\pm\rangle$ eigenstates in the first column correspond to the $|\pm\rangle$ states on the left-hand sides of Eqs. (13) and (14); the basis states in the second column correspond to the $\{|a\rangle, |b\rangle\}$ states on the right-hand sides of these equations. The entries of each 2×2 Hamiltonian, Eq. (10), are listed in the E_a, E_b, and V columns. The last three columns give specific names for the template energies in Eq. (11) and coefficients in Eq. (15). To calculate the eigenstates of a given row, substitute the values from the third through fifth columns into Eqs. (10)–(15).

$|\pm\rangle$	Basis states	E_a	E_b	V	$E_\pm$	η_a	η_b
$|l2;+;\pm\rangle$	$\{|pl2;+\rangle, |dl2;+\rangle\}$	$E_+^{(l)}$	$E_+^{(d)}$	$-V_{p,d}$	$E_{l,\pm}^{(+)}$	$\gamma_{p,l}^{(+)}$	$\gamma_{d,l}^{(+)}$
$|h2;+;\pm\rangle$	$\{|ph2;+\rangle, |dh2;+\rangle\}$	$E_+^{(l)}$	$E_+^{(d)}$	$-V_{p,d}$	$E_{l,\pm}^{(+)}$	$\gamma_{p,l}^{(+)}$	$\gamma_{d,l}^{(+)}$
$|so2;+;\pm\rangle$	$\{|pso2;+\rangle, |dso2;+\rangle\}$	$E_+^{(so)}$	$E_+^{(d)}$	$-V_{p,d}$	$E_{so,\pm}^{(+)}$	$\gamma_{p,so}^{(+)}$	$\gamma_{d,so}^{(+)}$
$|l2;-;\pm\rangle$	$\{|pl2;-\rangle, |dl2;-\rangle\}$	$E_-^{(l)}$	$E_-^{(d)}$	$-V_{p,d}$	$E_{l,\pm}^{(-)}$	$\gamma_{p,l}^{(-)}$	$\gamma_{d,l}^{(-)}$
$|h2;-;\pm\rangle$	$\{|ph2;-\rangle, |dh2;-\rangle\}$	$E_-^{(l)}$	$E_-^{(d)}$	$-V_{p,d}$	$E_{l,\pm}^{(-)}$	$\gamma_{p,l}^{(-)}$	$\gamma_{d,l}^{(-)}$
$|so2;-;\pm\rangle$	$\{|pso2;-\rangle, |dso2;-\rangle\}$	$E_-^{(so)}$	$E_-^{(d)}$	$-V_{p,d}$	$E_{so,\pm}^{(-)}$	$\gamma_{p,so}^{(-)}$	$\gamma_{d,so}^{(-)}$
$|\sigma2;+;\pm\rangle$	$\{|s\uparrow;+\rangle, |s^*\uparrow;+\rangle\}$	$E_+^{(s)}$	$E_+^{(s^*)}$	$-V_{s,s^*}$	$E_{\sigma,\pm}^{(+)}$	$\sigma_s^{(+)}$	$\sigma_{s*}^{(+)}$
$|\sigma2;-;\pm\rangle$	$\{|s\uparrow;-\rangle, |s^*\uparrow;-\rangle\}$	$E_-^{(s)}$	$E_-^{(s^*)}$	V_{s,s^*}	$E_{\sigma,\pm}^{(-)}$	$\sigma_s^{(-)}$	$\sigma_{s*}^{(-)}$

$$w_{v,so+-} = \frac{2}{E_{l,-}^{(-)} - E_{so,-}^{(+)}} \left(\frac{2}{3}\right) [\gamma_{d,l}^{(-)} \gamma_{d,so}^{(+)} V_{x,y} + \gamma_{p,l}^{(-)} \gamma_{p,so}^{(+)} U_{d,d}$$
$$+ (\gamma_{d,l}^{(-)} \gamma_{p,so}^{(+)} - \gamma_{p,l}^{(-)} \gamma_{d,so}^{(+)}) U_{p,d}]^2. \tag{23}$$

The other light valence band couplings are to the s-like (at zone center) bands,

$$w_{v,\sigma++} = \frac{2}{E_{l,-}^{(-)} - E_{\sigma,+}^{(+)}} \left(\frac{2}{3}\right) [-\gamma_{d,l}^{(-)} \sigma_s^{(+)} V_{s,p} - \gamma_{d,l}^{(-)} \sigma_{s*}^{(+)} V_{s*,p}$$
$$+ \gamma_{p,l}^{(-)} \sigma_s^{(+)} V_{s,d} + \gamma_{p,l}^{(-)} \sigma_{s*}^{(+)} V_{s*,d}]^2, \tag{24}$$

$$w_{v,\sigma+-} = \frac{2}{E_{l,-}^{(-)} - E_{\sigma,-}^{(+)}} \left(\frac{2}{3}\right) [\gamma_{d,l}^{(-)} \sigma_{s*}^{(+)} V_{s,p} - \gamma_{d,l}^{(-)} \sigma_s^{(+)} V_{s*,p}$$
$$+ \gamma_{p,l}^{(-)} \sigma_s^{(+)} V_{s*,d} - \gamma_{p,l}^{(-)} \sigma_{s*}^{(+)} V_{s,d}]^2, \tag{25}$$

as well as to the d-like $(3z^2 - r^2)$ band:

$$w_{v,\delta-} = \frac{2}{E_{l,-}^{(-)} - E_{\delta,-}} \left(\frac{2}{3}\right) [\gamma_{d,l}^{(-)} W_{p,d} + \gamma_{p,l}^{(-)} W_{d,d}]^2. \tag{26}$$

The split-off valence band couples to the heavy conduction bands,

$$w_{so,h++} = \frac{2}{E_{so,-}^{(-)} - E_{l,+}^{(+)}} \left(\frac{2}{3}\right) [-\gamma_{d,so}^{(-)} \gamma_{p,l}^{(+)} V_{x,y}$$
$$+ \gamma_{p,so}^{(-)} \gamma_{d,l}^{(+)} U_{d,d} + (\gamma_{d,so}^{(-)} \gamma_{d,l}^{(+)} + \gamma_{p,so}^{(-)} \gamma_{p,l}^{(+)}) U_{p,d}]^2, \tag{27}$$

$$w_{so,h+-} = \frac{2}{E_{so,-}^{(-)} - E_{l,-}^{(+)}} \left(\frac{2}{3}\right) [\gamma_{d,so}^{(-)} \gamma_{d,l}^{(+)} V_{x,y} + \gamma_{p,so}^{(-)} \gamma_{p,l}^{(+)} U_{d,d}$$
$$+ (\gamma_{d,so}^{(-)} \gamma_{p,l}^{(+)} - \gamma_{p,so}^{(-)} \gamma_{d,l}^{(+)}) U_{p,d}]^2, \tag{28}$$

the s-like bands,

$$w_{so,\sigma++} = \frac{2}{E_{so,-}^{(-)} - E_{\sigma,+}^{(+)}} \left(\frac{1}{3}\right) [-\gamma_{d,so}^{(-)} \sigma_s^{(+)} V_{s,p}$$
$$- \gamma_{d,so}^{(-)} \sigma_{s*}^{(+)} V_{s*,p} + \gamma_{p,so}^{(-)} \sigma_s^{(+)} V_{s,d}$$
$$+ \gamma_{p,so}^{(-)} \sigma_{s*}^{(+)} V_{s*,d}]^2, \tag{29}$$

$$w_{so,\sigma+-} = \frac{2}{E_{so,-}^{(-)} - E_{\sigma,-}^{(+)}} \left(\frac{1}{3}\right) [\gamma_{d,so}^{(-)} \sigma_{s*}^{(+)} V_{s,p} - \gamma_{d,so}^{(-)} \sigma_s^{(+)} V_{s*,p}$$
$$+ \gamma_{p,so}^{(-)} \sigma_s^{(+)} V_{s*,d} - \gamma_{p,so}^{(-)} \sigma_{s*}^{(+)} V_{s,d}]^2, \tag{30}$$

and the d-like band

$$w_{so,\delta-} = \frac{2}{E_{so,-}^{(-)} - E_{\delta,-}} \left(\frac{1}{3}\right) [\gamma_{d,so}^{(-)} W_{p,d} + \gamma_{p,so}^{(-)} W_{d,d}]^2. \tag{31}$$

With the above incompleteness corrections and couplings the [001] inverse valence band effective-mass formulas for the heavy (hh), light (lh), and split-off (soh) holes can be written as

$$\frac{m_0}{m_{[001]}^{(hh)}} = \frac{m_0}{\hbar^2} \left(\frac{a}{4}\right)^2 [w_{(inc),l}^{[001]} + w_{v,l++} + w_{v,l+-} + w_{v,so++}$$
$$+ w_{v,so+-}], \tag{32}$$

$$\frac{m_0}{m_{[001]}^{(lh)}} = \frac{m_0}{\hbar^2} \left(\frac{a}{4}\right)^2 [w_{(inc),l}^{[001]} + w_{v,l++} + w_{v,l+-} + w_{v,\sigma++}$$
$$+ w_{v,\sigma+-} + w_{v,\delta-}], \tag{33}$$

$$\frac{m_0}{m^{(soh)}} = \frac{m_0}{\hbar^2} \left(\frac{a}{4}\right)^2 [w_{(inc),so}^{[001]} + w_{so,l++} + w_{so,l+-} + w_{so,\sigma++}$$
$$+ w_{so,\sigma+-} + w_{so,\delta-}], \tag{34}$$

where a is the conventional unit cell cube edge and m_0 is the free-electron mass. Next we consider the heavy and light valence band masses along [111] and [110].

C. Inverse heavy and light masses along [111] and [110]

The task of finding these inverse mass formulas is rather more complicated than along [001] since for **k** along [111] the matrix $d\underline{\mathbf{H}}/dk$ mixes the 1 and 2 states, and along either direction it mixes the heavy and light holes; the matrix $d^2\underline{\mathbf{H}}/dk^2$ also complicates matters. The inverse mass matrix[15] of dimension either 4×4 along [111] or 2×2 along [110] must be diagonalized. The [111] calculation is the more easily performed despite its larger dimensionality and the results are presented first.

In the second-near-neighbor, spin-orbit $sp^3 s*$ model, it is known that the [111] inverse mass matrix in the basis $\{|l1; -;-\rangle, |h1;-;-\rangle, |l2;-;-\rangle, |h2;-;-\rangle\}$ takes the form[8]

$$\underline{\mathbf{W}} = \begin{bmatrix} \mu & i\tau & 0 & (1-i)\tau \\ -i\tau & \mu & (-1+i)\tau & 0 \\ 0 & (-1-i)\tau & \mu & -i\tau \\ (1+i)\tau & 0 & i\tau & \mu \end{bmatrix},$$
$$\mu, \tau \in \mathrm{Re}, \tag{35}$$

which has eigenvectors independent of μ and τ. Since the spin-orbit coupling in the nearest-neighbor $sp^3 d^5 s*$ model is the same as in the $sp^3 s*$ model, one expects that these very eigenvectors will diagonalize the [111] inverse mass matrix in the $sp^3 d^5 s*$ model also. This is in fact the case. Taking advantage of the fact that these states diagonalize the inverse mass matrix allows calculation of the masses as in Sec. II B above. Here an additional incompleteness correction is found,

$$\Delta w_{(inc),l}^{[111]} = \frac{2}{3} [(\gamma_{d,l}^{(-)})^2 V_{x,y} - (\gamma_{p,l}^{(-)})^2 U_{d,d} - 2\gamma_{d,l}^{(-)} \gamma_{p,l}^{(-)} U_{p,d}], \tag{36}$$

which is added to the light inverse mass and subtracted from the heavy inverse mass. All of the interband terms have been calculated above; they are merely allocated differently between the heavy and light masses. The complete formulas are thus

$$\frac{m_0}{m_{[111]}^{(hh)}} = \frac{m_0}{\hbar^2}\left(\frac{a}{4}\right)^2\left[w_{(inc),l}^{[001]} - \Delta w_{(inc),l}^{[111]} + w_{v,l++} + w_{v,l+-}\right.$$
$$\left. + \frac{3}{4}w_{v,\delta-}\right], \qquad (37)$$

$$\frac{m_0}{m_{[111]}^{(lh)}} = \frac{m_0}{\hbar^2}\left(\frac{a}{4}\right)^2\left[w_{(inc),l}^{[001]} + \Delta w_{(inc),l}^{[111]} + w_{v,l++} + w_{v,l+-}\right.$$
$$+ w_{v,so++} + w_{v,so+-} + w_{v,\sigma++} + w_{v,\sigma+-}$$
$$\left. + \frac{1}{4}w_{v,\delta-}\right]. \qquad (38)$$

From Eqs. (37) and (38) it is apparent that the inverse masses obey the sum rule

$$\frac{m_0}{m_{[111]}^{(lh)}} + \frac{m_0}{m_{[111]}^{(hh)}} = \frac{m_0}{m_{[001]}^{(lh)}} + \frac{m_0}{m_{[001]}^{(hh)}}. \qquad (39)$$

At this point we have all the ingredients needed to calculate the [110] masses.

Along [110] the 1 and 2 states are not mixed but the heavy and light states are, so that there are two two-dimensional inverse mass matrices. Since one is the complex conjugate of the other, only one needs to be diagonalized. For the 1 states in the basis $\{|l1;-;-\rangle,|h1;-;-\rangle\}$, this matrix is most usefully written

$$\mathbf{W}_{[110]}^{(1)} = \begin{bmatrix} \bar{w}_{[001]} - \frac{1}{4}\delta_{[001]} & -i\frac{\sqrt{3}}{4}\delta_{[111]} \\ i\frac{\sqrt{3}}{4}\delta_{[111]} & \bar{w}_{[001]} + \frac{1}{4}\delta_{[001]} \end{bmatrix}, \qquad (40)$$

where

$$\bar{w}_{[001]} = \frac{1}{2}\left[\frac{1}{m_{[001]}^{(lh)}} + \frac{1}{m_{[001]}^{(hh)}}\right], \qquad (41)$$

$$\delta_{[001]} = \frac{1}{m_{[001]}^{(lh)}} - \frac{1}{m_{[001]}^{(hh)}}, \qquad (42)$$

$$\delta_{[111]} = \frac{1}{m_{[111]}^{(lh)}} - \frac{1}{m_{[111]}^{(hh)}}. \qquad (43)$$

Remarkably, for this model with d orbitals, the [110] inverse masses are functions solely of the [001] and [111] inverse masses: this is the same as in the sp^3s^* model.[8]

The eigenvalues of the inverse mass matrix give the inverse effective masses. Diagonalizing Eq. (40) one finds

$$\frac{m_0}{m_{[110]}^{(lh)}} = \frac{1}{2}\left[\frac{m_0}{m_{[001]}^{(lh)}} + \frac{m_0}{m_{[001]}^{(hh)}}\right]$$
$$- \frac{1}{4}\sqrt{\left[\frac{m_0}{m_{[001]}^{(lh)}} - \frac{m_0}{m_{[001]}^{(hh)}}\right]^2 + 3\left[\frac{m_0}{m_{[111]}^{(lh)}} - \frac{m_0}{m_{[111]}^{(hh)}}\right]^2}, \qquad (44)$$

TABLE IV. Tight-binding parameters for Si and Ge (same-site and two-center integrals) in the Slater-Koster notation (Ref. 3); units are eV.

Parameter	Si	Ge
E_s	-2.15168	-1.95617
E_p	4.22925	5.30970
E_{s*}	19.11650	19.29600
E_d	13.78950	13.58060
λ	0.01989	0.10132
$ss\sigma$	-1.95933	-1.39456
$s*s*\sigma$	-4.24135	-3.56680
$ss*\sigma$	-1.52230	-2.01830
$sp\sigma$	3.02562	2.73135
$s*p\sigma$	3.15565	2.68638
$sd\sigma$	-2.28485	-2.64779
$s*d\sigma$	-0.80993	-1.12312
$pp\sigma$	4.10364	4.28921
$pp\pi$	-1.51801	-1.73707
$pd\sigma$	-1.35554	-2.00115
$pd\pi$	2.38479	2.10953
$dd\sigma$	-1.68136	-1.32941
$dd\pi$	2.58880	2.56261
$dd\delta$	-1.81400	-1.95120

$$\frac{m_0}{m_{[110]}^{(hh)}} = \frac{1}{2}\left[\frac{m_0}{m_{[001]}^{(lh)}} + \frac{m_0}{m_{[001]}^{(hh)}}\right]$$
$$+ \frac{1}{4}\sqrt{\left[\frac{m_0}{m_{[001]}^{(lh)}} - \frac{m_0}{m_{[001]}^{(hh)}}\right]^2 + 3\left[\frac{m_0}{m_{[111]}^{(lh)}} - \frac{m_0}{m_{[111]}^{(hh)}}\right]^2}. \qquad (45)$$

Note that the valence band masses are negative. Again, these results are the same as in the sp^3s^* model. The next section discusses the implications of these results for the reproducibility of the model.

III. RESULTS: Si AND Ge PARAMETERS

An important component of the NEMO theory and software development has been representing material properties such as effective masses and band gaps of the centrally relevant conduction and valence bands at room temperature[12,13] for the quantitative simulation of electronic structure and electronic transport in heterostructures. The low-temperature Si and Ge parametrizations of Jancu *et al.*[5] served as a starting point in the effort to develop the ability to model SiGe alloys and arbitrary distortions of the Si-Si and Ge-Ge bonds.[14] Of particular interest has been the representation of the material properties at room temperature, and where available experimental room temperature data[17] are used as a parametrization target. The parametrizations used were generated by an automated genetic algorithm approach detailed in Ref. 11, followed by hand tuning using the formulas presented in Sec. II above. Table IV lists our generated tight-binding parameters for Si and Ge, while Table V gives selected energy band edges, gaps, and effective masses

TABLE V. Selected energies (in eV), k minima (in relative units of the Brillouin zone edge), and effective masses (in units of the free-electron mass) for Si and Ge as reproduced by the parameters of Table IV, along with target values given our fitting algorithm. Unless otherwise noted, targets are from Ref. 17. In both materials the L-valley minima occur at the symmetry point. All masses are computed at the corresponding extrema. Values marked with an asterisk are not well established and as a result were not targeted strongly in the optimization.

Quantity	Si	Si target	% deviation	Ge	Ge target	% deviation
E_c^Γ	3.399	3.368	0.9	1.584	1.575	0.6
E_v^Γ	0.0	0.0	0.0	0.770	0.770	0.0
Δ_0	0.0472	0.045	4.9	0.225	0.300	25.0
$E_{c,min}^L$	2.383	2.400[a]	0.7	1.448	1.434	1.0
$E_{c,\min}^X$	1.131	1.118	1.2	1.676	1.675[b]	0.1
$k_{min}^{[001]}$	81.3%	85.0%	4.4	88.5%	85.0%	4.1
$m_{X,l}^{(e)}$	0.891	0.916	2.7	0.701	0.900*	22.1
$m_{X,t}^{(e)}$	0.201	0.190	5.8	0.201	0.200*	0.5
$m_{L,l}^{(e)}$	3.433	2.000	71.7*	1.584	1.590	0.4
$m_{L,t}^{(e)}$	0.174	0.100	74.0*	0.0813	0.0823	1.2
$m_{lh}^{[001]}$	-0.214	-0.204	4.9	-0.0488	-0.0457	6.8
$m_{lh}^{[110]}$	-0.152	-0.147	3.4	-0.0424	-0.0417	1.7
$m_{lh}^{[111]}$	-0.144	-0.139	3.6	-0.0410	-0.0406	1.0
$m_{hh}^{[001]}$	-0.276	-0.275	0.4	-0.173	-0.211	18.0
$m_{hh}^{[110]}$	-0.581	-0.579	0.3	-0.368	-0.383	3.9
$m_{hh}^{[111]}$	-0.734	-0.738	0.5	-0.531	-0.502	5.8
m_{so}	-0.246	-0.234	5.1	-0.0947	-0.095	0.3

[a]Reference 18.
[b]See Ref. 22.

reproduced by them. Care was taken to simultaneously match the critical band edges at Γ, X,L, and X_{min} on the Δ line as well as the various anisotropic conduction band masses at L and X_{min} and the anisotropic valence band masses. The present parameters place the Si L valley higher than do those of Ref. 5, since experimental evidence (Ref. 18) indicates that it should be higher (2.40 eV). Little weight was assigned to the importance of the conduction band mass at Γ and the split-off band effective mass; the split-off valence band splitting Δ_0 was also assigned a fairly small weight. In Ge the conduction band effective masses at X_{min} were optimized with a relatively small importance weight. In Si the conduction band effective masses at L were not optimized at all. The relative errors in percent are listed in Table V and indicate values typically less than 5%. The sole exception is the Ge heavy-hole mass along [001], which, as discussed below, allows little freedom in fitting. (Note as well that Ref. 5 achieves a fairly small value for this mass, -0.196.) We consider this accuracy acceptable within the range of known experimental data.

The resulting bands of Si and Ge in an energy range around the valence and lowest conduction bands are plotted in Figs. 1 and 2. Note that the valence bands of Ge are offset with respect to those of Si to facilitate their use in future Si/Ge heterostructure calculations.

One feature of the bands at Γ that immediately stands out is the difference in the conduction band structures of the two materials. Note, in particular, that the lowest conduction bands in Si (Fig. 1) are all nearly degenerate, whereas in Ge

(Fig. 2) the lowest conduction band is much nearer the valence bands than any of the others. The form of the inverse mass formulas suggests that this will have profound consequences on the masses reproduced by the model for the two materials, and this is indeed the case.

IV. DISCUSSION

It is instructive to compare the capabilities of the nearest-neighbor sp^3d^5s* model to those of the widely employed
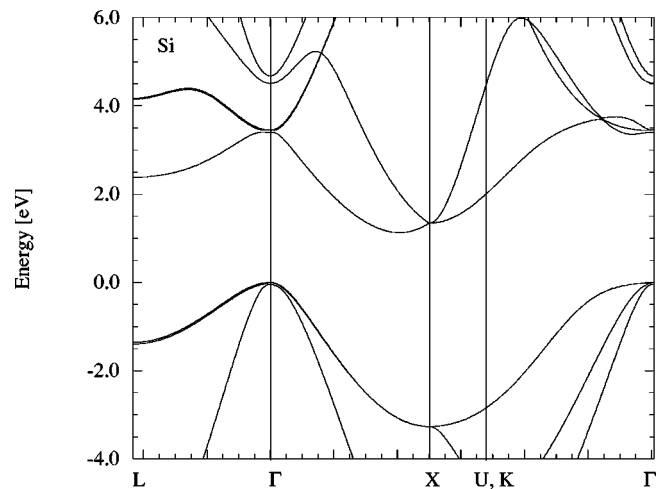

FIG. 1. Bands of silicon as reproduced by our parameters given in Table IV; only the central bands are shown.
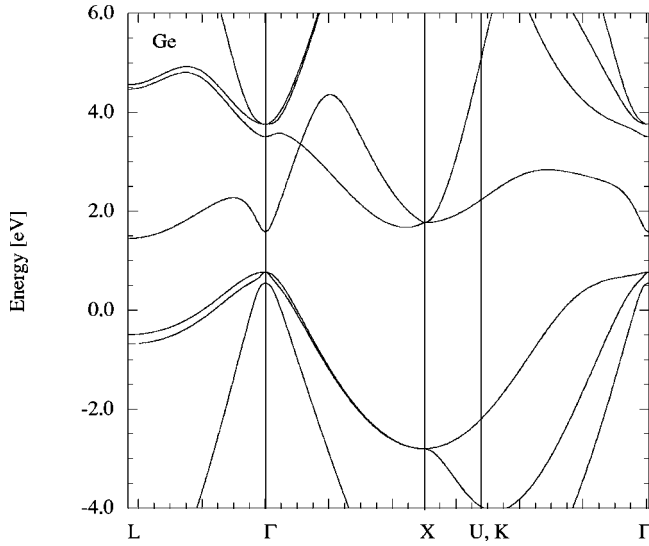
FIG. 2. Bands of germanium as reproduced by our parameters given in Table IV; only the central bands are shown.

sp^3s^* model. The number of parameters needed for the nearest-neighbor $sp^3d^5s^*$ model is comparable to that of the second-near-neighbor sp^3s^* model and of course significantly more than that of the nearest-neighbor version. Reference 11 discusses in detail the limitations of the nearest-neighbor sp^3s^* model. It can be shown analytically for the nearest-neighbor sp^3s^* model that at the X-point face the transverse effective mass is infinitely large, i.e., the bands in that direction are flat. Furthermore, Ref. 11 shows that the conduction band minimum on the Δ line must be pushed about 60% of the way to the Brillouin zone face to be able to achieve a transverse mass small enough to give the conduction band isosurfaces their desired cigarlike shape. Even so, the mass remains about 50% too large, and the improper position of the conduction band minimum at about 60% rather than 85% results in improperly quantized wave functions[19] in heterostructures and incorrect reproduction of the valley splitting.[19] Good longitudinal and transverse conduction band masses come at the price of severe distortion of the valence bands, where the masses are much too small (by factors of 10–20), and very poor reproduction of the higher conduction bands. A simultaneous fit of the conduction and valence bands in the nearest-neighbor sp^3s^* model for Si is simply impossible,[11] rendering the sp^3s^* nearest-neighbor model useless for realistic calculations for heterostructures in which conduction and valence bands couple and even for simple devices such as Zener diodes which depend on interband tunneling. In contrast, Ref. 11 shows that the second-nearest-neighbor sp^3s^* model can simultaneously reproduce the valence and conduction band effective masses and relevant band gaps.

Together, Ref. 11 and this work show that the second-nearest-neighbor sp^3s^* and the nearest-neighbor $sp^3d^5s^*$ models are comparable in their ability to simultaneously reproduce valence and conduction band properties. In terms of their use for transport simulations[13] in NEMO 1D both models require the eigenvalues of a 40×40 matrix for the computation of open system boundary conditions. For a 1D (zinc blende) heterostructure which is spatially resolved in N anion and N cation layers, the $sp^3d^5s^*$ model requires twice as much computational effort compared to the sp^3s^* second-nearest-neighbor model, since the Hamiltonian needs to be resolved in the recursive Green function (RGF) algorithm as $2N\times(20\times20)$ blocks compared to $N\times(20\times20)$ blocks. Since the RGF scales as $M\times B^3$, where B is the block size and M the number of blocks, the second-nearest-neighbor sp^3s^* model runs twice as fast as the nearest-neighbor $sp^3d^5s^*$ model.[20]

For 3D nanoelectronic structures, however, which may contain arbitrary lattice distortions, the $sp^3d^5s^*$ model provides significant advantages over the second-nearest-neighbor sp^3s^* model. The nearest-neighbor model allows for a well-defined scaling of the on-site and nearest-neighbor interactions[14] due to strain-induced distortions. The second-nearest-neighbor model requires a much more complicated treatment of the second-nearest-neighbor interaction parameters, due to the more complicated three-center integrals. Furthermore, the nearest-neighbor model makes the simulation of disordered materials such as alloys and interfaces much more straightforward, since at these inhomogeneities the physics behind the nearest-neighbor interpolation rules is solid. For second-near-neighbor interactions, however, the interpolation rules are far less certain. These fundamental advantages of the nearest-neighbor model have led the NEMO 3D development[12] to focus on the $sp^3d^5s^*$ model, rather than the second-near-neighbor sp^3s^* model; the NEMO 1D tool supports both models.

In the automated fit of the $sp^3d^5s^*$ model we found that achieving good results for the heavy- and light-hole masses in all three directions ([001], [110], and [111]) was much more difficult (in terms of the computer time required) for Ge than for Si. This is due in large measure to the very different conduction band structures of these two materials. In Si, the three lowest conduction bands (p- and d-like) are very close, all lying around 3.4 eV above the valence band maximum; the lowest s-like conduction band is only another 1.1 eV higher. In contrast, the lowest conduction band of Ge is s-like and only about 0.8 eV above the valence band maximum; the next highest (p- and d-like) conduction bands are approximately another 2 eV higher. Since the hole masses derived in Secs. II B and II C above are largely determined by the couplings between the valence and conduction bands, the conduction band ordering has a major impact on the ease with which the model can reproduce them; a detailed look at the components of the various inverse mass formulas reveals how the different conduction band structure affects the masses in the two materials.

First consider the light hole masses. Table VI lists the (dimensionless) interband couplings and their corresponding energy gaps; also listed are the incompleteness term for [001] and the incompleteness difference term for [111]. As one might expect, the p- and d-like valence bands couple strongly to the s-like conduction bands, especially the lowest. In Ge this one coupling dominates all other terms, since its energy gap (recall that each w is an energy coupling squared divided by an energy gap) is less than one-third of the next smallest gap. Conversely, in Si, the two largest cou-

TABLE VI. Couplings (dimensionless) and energy gaps (in eV) of valence bands to various conduction bands for Ge and Si parameters given in Table IV. Also listed are the dimensionless incompleteness correction term for [001] and the incompletness correction difference for [111]; see Eqs. (18) and (36).

	Ge		Si	
	$\frac{m_0}{\hbar^2}\left(\frac{a}{4}\right)^2 w$	E_{gap}	$\frac{m_0}{\hbar^2}\left(\frac{a}{4}\right)^2 w$	E_{gap}
$v,\sigma+-$	-19.118	0.814	-3.588	4.511
$v,so+-$	-4.821	2.739	-3.268	3.399
$v,l+-$	-2.184	2.990	-1.608	3.450
$v,\sigma++$	-0.734	34.831	-0.676	37.256
$v,so++$	-0.373	18.459	-0.175	19.169
$v,l++$	-0.190	18.512	-0.088	19.178
$v,\delta-$	-0.063	8.579	-0.225	9.305
inc: l, [001]	1.787		1.513	
inc: l, [111]	1.249		1.012	

plings are roughly the same size,[21] as are the three lowest energy gaps. In Si, therefore, three terms strongly influence the light-hole mass as compared to a single term in Ge. As a result, the model does not permit the Ge light-hole mass to change much in going from [001] to [111], while it tends to change that of Si relatively more due to the transfer of the valence band couplings to the split-off conduction bands from the heavy hole along [001] to the light hole along [111]—compare Eqs. (32) and (38). This observation is reflected both experimentally and theoretically (Table V) where in Si $m_{lh}^{[111]} \approx 0.67 m_{lh}^{[001]}$, while for Ge $m_{lh}^{[111]} \approx 0.89 m_{lh}^{[001]}$; the absolute change in Si is almost 14 times that in Ge. Even though the experimental data indicate a relatively small change in the Ge light-hole mass, the fact that this mass is dominated by a single term makes fitting the experimental values significantly harder in Ge than in Si.

The conduction band structure of the two materials also affects the heavy-hole masses. Table V shows that in both materials the heavy-hole mass increases significantly in magnitude on going from [001] to [111], and the change is smaller in both relative and absolute terms for Ge than for Si. The model reflects and to some extent overemphasizes this trend (Table VI), since the largest component of the [001] heavy-hole mass is the coupling to the lowest split-off conduction band, which appears in the [001] mass but not in the [111] mass. This very large change in mass dictated by the model is therefore more compatible with Si than with Ge, and thus explains the difficulty in fitting the Ge heavy-hole masses along [001] and [111]. Thus these formulas show the important connection between the conduction band position and ordering and the valence band masses along [001] and [111].

The analytical mass equations detail clearly that in spite of the inclusion of d orbitals the [110] masses are functions solely of those along [001] and [111]. This result is the same as one finds in the sp^3s^* models or indeed in $\mathbf{k} \cdot \mathbf{p}$ perturbation theory based on s- and p-like orbitals. (Note that in the present model the spin-orbit coupling is modeled by a same-atom, p-orbital-only parameter.) Thus, where precise experimental results on effective masses are available, the suitability of this model for diamond semiconductors can be checked with the sum rule Eq. (39) and the inverse mass relations Eqs. (44) and (45).

V. CONCLUSIONS

Exact, analytic formulas for the inverse valence band effective masses along [001], [110], and [111] for the spin-orbit, nearest-neighbor $sp^3d^5s^*$ Hamiltonian[5] for diamond at Γ have been derived. These formulas are helpful in manual or automated fits of the bands. These results show the relationships between the valence band masses along [001], [110], and [111], as well as the intimate connection between these masses and the ordering and position of the various conduction bands. This connection has an interesting corollary for automated fitting procedures: The connection between the conduction band position and ordering and the valence band masses along [001], [110], and [111] means that fitting both energy gaps and curvatures will help ensure (but cannot guarantee since no model is perfectly complete) proper ordering of the bands at the zone center. A parametrization of the $sp^3d^5s^*$ model for unstrained Si and Ge is presented. In the parametrization particular care was given to the reproduction of the lowest conduction bands and upper valence bands which are relevant to semiconductor heterostructure modeling, and deviations from the desired target values of typically less than 5% were found.

*Present address: School of Electrical & Computer Engineering, Purdue University, West Lafayette, Indiana 47907.

[1] M. Friesen *et al.*, Phys. Rev. B **67**, 121301 (2003).

[2] Timothy B. Boykin, Gerhard Klimeck, Mark Eriksson, Mark Friesen, S. N. Coppersmith, Paul von Allmen, Fabiano Oyafuso, and Seungwon Lee (unpublished).

[3] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).

[4] P. Vogl, Harold P. Hjalmarson, and John D. Dow, J. Phys. Chem. Solids **44**, 365 (1983).

[5] Jean-Marc Jancu, Reinhold Scholz, Fabio Beltram, and Franco Bassani, Phys. Rev. B **57**, 6493 (1998).

[6] Timothy B. Boykin, Gerhard Klimeck, R. Chris Bowen, and Roger Lake, Phys. Rev. B **56**, 4102 (1997).

[7] Timothy B. Boykin, Phys. Rev. B **56**, 9613 (1997).

[8] Timothy B. Boykin, Lisa J. Gamble, R. Chris Bowen, and Gerhard Klimeck, Phys. Rev. B **59**, 7301 (1999).

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

[10] D. Helmholz and L. C. Lew Yan Voon, Phys. Rev. B **65**, 233204 (2002).

[11] Gerhard Klimeck, R. Chris Bowen, Timothy B. Boykin, Carlos Salazar-Lazaro, Thomas A. Cwik, and Adrian Stoica, Superlattices Microstruct. **27**, 77 (2000); Gerhard Klimeck, R. Chris Bowen, Timothy B. Boykin, and Thomas A. Cwik, *ibid.* **27**, 519 (2000).

[12] Gerhard Klimeck, Fabiano Oyafuso, Timothy B. Boykin, R. Chris Bowen, and Paul von Allmen, Comput. Model. Eng. Sci. **3**, 601 (2002).

[13] Roger Lake, Gerhard Klimeck, R. Chris Bowen, and Dejan Jovanovic, J. Appl. Phys. **81**, 7845 (1997).

[14] Timothy B. Boykin, Gerhard Klimeck, R. Chris Bowen, and Fabiano Oyafuso, Phys. Rev. B **66**, 125207 (2002).

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

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

[17] Available experimental data (unless otherwise noted) are taken from *Semiconductors: Group IV Elements and III-V Compounds*, edited by O. Madelung (Springer, New York, 1991).

[18] D. Straub, L. Ley, and F. J. Himpsel, Phys. Rev. Lett. **54**, 142 (1985).

[19] T. B. Boykin, G. Klimeck, M. Eriksson, M. Friesen, S. N. Coppersmith, P. von Allmen, F. Oyafuso, and S. Lee, Appl. Phys. Lett. **84**, 115 (2004). During the course of this research, we tried different Si parametrizations and found that those with an X-valley minimum significantly different from 85% produced incorrect results. Reference 11 lists as an optimization target for the second-near-neighbor sp^3s^* model an X-valley minimum at 75% of the way to the Brillouin zone face, which is a little too small; we have been able to adjust the parameters of that model to achieve a minimum at 85%.

[20] Both models produce block-tridiagonal Hamiltonian matrices. The nearest-neighbor $sp^3d^5s^*$ model couples adjacent atomic planes, so that for 20 orbitals per atom the block size is 20, giving a Hamiltonian for N anion and N cation planes of size $2N \times (20 \times 20)$. The second-near-neighbor sp^3s^* model couples adjacent monolayers (one anion and one cation plane). Here there are ten orbitals per atom, and two atomic planes per monolayer, so that the block size is once again 20. Since there are N monolayers, the dimension of the Hamiltonian matrix is now only $N \times (20 \times 20)$. The RGF algorithm partitions the open-system Hamiltonian for a planar device into one active region with a contact region on each end. To decouple an active (or quantum) region of N layers (represented by block matrices of dimension B) from the full Hamiltonian, the RGF algorithm requires one step per block, working from one contact to the other. Each step in this forward part of the algorithm involves matrix inversions and multiplications so the operation count is of order $N \times B^3$. To obtain charge densities or the density of states involves back substitutions, so that these calculations scale as $N \times B^2$.

[21] The coupling to the lowest s-like conduction band is a little larger than any other even though the energy gap is a little larger than the other gaps to low conduction bands.

[22] The locations of the Ge X-valley minima do not seem well established. Rubin Braunstein, Arnold R. Moore, and Frank Herman, Phys. Rev. **109**, 695 (1958) estimate that the Ge X-valley minima lie about 0.22 eV above the L-valley minima at room temperature, based on extrapolation of their experimental energy gap curves for $Si_{1-x}Ge_x$ alloys. We opt for a slightly larger separation, about 0.24 eV.