Strain-induced, off-diagonal, same-atom parameters in empirical tight-binding theory suitable for [110] uniaxial strain applied to a silicon parametrization

Timothy B. Boykin

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

Mathieu Luisier, Mehdi Salmani-Jelodar, and Gerhard Klimeck

Network for Computational Nanotechnology, Birk Nanotechnology Center, School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, USA

State-of-the-art transistors achieve their improved performance through strain engineering. The somewhat unusual uniaxial [110] strain is of particular importance as it provides a significant mobility increase for electrons. Empirical tight binding has shown tremendous benefits in modeling realistically large structures including standard strain conditions, but often fails to predict the correct uniaxial [110] strain behavior because most treatments neglect the same-atom different-orbital matrix elements induced by this strain. Two separate mechanisms are responsible for these conditions: Löwdin orbital changes and displacement of nearest-neighbor potentials. We present a model which separately includes both mechanisms via parameters whose range of validity can be independently determined. Using this method we optimize a set of strain parameters for Si. The combination of both effects is able to reproduce the Si X_z-valley transverse mass splitting under uniaxial [110] strain. We then use this model to calculate the drain current of a strained double-gate, ultrathin-body metal-oxide-semiconductor field-effect transistor, finding experimentally plausible results.

I. INTRODUCTION

Ab initio methods (e.g., the various forms of DFT) have proven highly accurate and useful in bulk and for very small atomic clusters. However, their computational burden for even the electronic structure of nanodevices is very high, and often becomes prohibitive when transport or optical effects must be included. Furthermore, most are plane wave based and like all such methods require large numbers of basis states to resolve sharp interfaces or alloy disorder. Consequently, computational experiments to determine the appropriate number of plane waves and cutoff must be performed for each particular geometry or structure.

Empirical tight binding, especially when parametrized using results from *ab initio* calculations or experiments, can be highly accurate and computationally efficient. The localized tight-binding basis, determined solely by the number of atoms and number of orbitals per atom, is well suited for modeling device or materials changes on a nanometer scale. Properly parametrized, empirical tight-binding models have demonstrated quantitative accuracy for resonant tunneling diodes,[1] quantum dots,[2] quantum wells,[3] and even single impurities.[4]

Over the past half-century since the pioneering work of Slater and Koster[5] many improvements have been made to the method. Vogl *et al.*[6] added an excited *s*-like orbital (s^*) to the traditional nearest-neighbor sp^3 basis to greatly improve the conduction bands of many diamond and zinc blende semiconductors. The addition of excited *d*-like orbitals by Jancu *et al.*[7] further improved nearest-neighbor models by making accurate reproduction of the *X*-valley transverse mass possible. (With an sp^3 or sp^3s^* basis a nearest-neighbor model cannot accurately reproduce the *X*-valley transverse mass and second- or more-distant-neighbor interactions are necessary.[8,9]) This improvement by Jancu *et al.*[7] significantly

aided nanostructure modeling because the handling of interfaces is much more transparent in nearest-neighbor models than in second- or more-distant-neighbor treatments, where considerable uncertainty attaches to the interface parameters.

Increasingly, strained nanodevices have become technologically relevant. Partly this increased relevance is due to scientific interest in quantum dot structures self-assembled from materials with different lattice constants (e.g., SiGe or InGaAs alloys or InAs/GaAs quantum dots). Of perhaps greater technological importance is the current drive for improved performance in modern Si transistors via strain engineering. Indeed, all current Si-based 65, 45, and 32 nm technology-node Si transistors utilize strain engineering to enhance electron and hole mobilities. In Si devices, there is now a particular interest in uniaxial [110]-strained devices because this strain splits the *X*-valley transverse masses, $m^*_{X,[110]}$ and $m^*_{X,[\bar{1}10]}$ for the two valleys along [001].[10,11] The interplay of strain, interfaces, and transport at the nanometer scale demands accurate treatment of strain in tight binding.

Strain has traditionally been incorporated into tight-binding models by scaling the two-center integrals using a generalization of Harrison's d^{-2} law:[12] $U = (d_0/d)^\eta U_0$, where U_0 is an ideal-crystal two-center integral, U its counterpart in the strained crystal, d_0 the ideal bond length, d the strained bond length, and η the scaling exponent. More recently, diagonal shifts for the *p* (Ref. 13) or *d* orbitals[7] or to all orbitals generally[14] have been included to improve the tight-binding reproduction of strain behavior. The case of uniaxial [110] strain, however, is not well reproduced with these approaches, largely because it induces same-atom, different-orbital Hamiltonian matrix elements. Niquet *et al.*[15] have addressed this problem by including effects on both diagonal and off-diagonal same-atom Hamiltonian matrix elements due to changes in the nearest-neighbor potentials. As dis-

cussed in Ref. 15, their model addresses both displacement of the nearest-neighbor potentials and orbital renormalization effects (necessarily including Löwdin orbital[16] renormalization as the method is orthogonal). While they are able to achieve a good fit of the Si and Ge strain behavior to *ab initio* results, this agreement comes at the cost of several of their β-strain parameters having signs opposite to what one expects (Sec. IVA of Ref. 15). The lack of clear physical guidance on the strain parameters is problematic because physically questionable parameter sets, which successfully reproduce bulk bands often do not give good nanostructure results. In a model which lumps all effects into the same parameters determining whether or not parameter values are reasonable is difficult or impossible. This problem can be circumvented by modeling the strain in such a way that changes in the nearest-neighbor potentials and Löwdin orbital[16] effects are clearly separated. Separating these effects offers more insight into the parameter meanings and better control of the parameter values.

Thus a model which clearly separates the effects of nearest-neighbor potentials and Löwdin orbital[16] renormalization is the aim of the present work. We generalize our earlier result[14] based on Löwdin orbital[16] changes to include off-diagonal, same-atom matrix elements. The more general treatment of the Löwdin procedure[16] incorporates changes to the matrix elements due to the nearest-neighbor potentials in a completely transparent manner, separating the orbital and potential effects, and allowing reasonable limits to be placed on both. Section II presents our model; Sec. III our results, including both bulk and application to a strained metal-oxide-semiconductor field-effect transistor (MOSFET); and Sec. IV our conclusions.

II. MODEL

A. Löwdin orbital matrix elements

Here we generalize our previous results on diagonal parameter shifts.[14] Because this derivation parallels the earlier

result, we briefly sketch the method, emphasizing the relaxation of several assumptions. As before, we scale the two-center integrals using the generalized form of Harrison's[12] law

$$U = U_0 \left(\frac{d_0}{d} \right)^{\eta}, \quad (1)$$

where U_0 is an ideal-crystal two-center integral (e.g., $V_{sp\sigma}$, $V_{pp\pi}$), U its counterpart in the strained crystal, d_0 the ideal bond length, d the strained bond length, and η the scaling exponent. The Löwdin procedure[16] gives the expression for the orthogonal atomiclike orbital of type α on atom i, $|\varphi_{i,\alpha}\rangle$ in terms of true atomic orbitals $|\Phi_{i,\alpha}\rangle$. To lowest order in the overlaps $S_{(j,\beta),(i,\alpha)}$, the result[16] is

$$|\varphi_{i,\alpha}\rangle \approx |\Phi_{i,\alpha}\rangle - \frac{1}{2} \sum_{j,\beta} S_{(j,\beta),(i,\alpha)} |\Phi_{j,\beta}\rangle, \quad (2)$$

$$S_{(j,\beta),(i,\alpha)} \approx \langle \Phi_{j,\beta} | \Phi_{i,\alpha} \rangle - \delta_{i,j} \delta_{\alpha,\beta}, \quad (3)$$

where we also assume that atomic orbitals on the same atom are orthonormal, $\langle \Phi_{i,\beta} | \Phi_{i,\alpha} \rangle = \delta_{\alpha,\beta}$. We denote the diagonal Löwdin matrix elements by $\varepsilon_{i,\alpha} = \langle \varphi_{i,\alpha} | \hat{H} | \varphi_{i,\alpha} \rangle$ and the corresponding diagonal atomic-orbital matrix elements as: $\varepsilon_{i,\alpha}^{(0)} = \langle \Phi_{i,\alpha} | \hat{H} | \Phi_{i,\alpha} \rangle$. As before we generalize the result of Wills and Harrison[17] taking the overlaps to be

$$S_{(n',\mu'),(n,\mu)} = K_{(n',\mu'),(n,\mu)} \frac{v_{(n',\mu'),(n,\mu)}}{\varepsilon_{(n',\mu')}^{(0)} + \varepsilon_{(n,\mu)}^{(0)}}, \quad n' \neq n, \quad (4)$$

where the $K_{(n',\mu'),(n,\mu)}$ are fitting parameters and $v_{(n',\mu'),(n,\mu)}$ a Hamiltonian matrix element between Löwdin orbitals of types μ' and μ on atoms n' and n. Carrying out the analysis as before,[14] but here for two arbitrary orbitals on the same atom, we find for the Löwdin matrix element

$$\langle \varphi_{i,\gamma} | \hat{H} | \varphi_{i,\alpha} \rangle \approx \langle \Phi_{i,\gamma} | \hat{H} | \Phi_{i,\alpha} \rangle - \frac{1}{2} \sum_{j \in \text{NN to } i, \beta} v_{(i,\gamma),(j,\beta)} v_{(j,\beta),(i,\alpha)} \left[\frac{K_{(j,\beta),(i,\alpha)}}{\varepsilon_{j,\beta}^{(0)} + \varepsilon_{i,\alpha}^{(0)}} + \frac{K_{(i,\gamma),(j,\beta)}}{\varepsilon_{j,\beta}^{(0)} + \varepsilon_{i,\gamma}^{(0)}} \right]$$
$$- \frac{1}{4} \sum_{j \in \text{NN to } i, \beta} v_{(i,\gamma),(j,\beta)} v_{(j,\beta),(i,\alpha)} K_{(i,\gamma),(j,\beta)} K_{(j,\beta),(i,\alpha)} \left[\frac{1}{\varepsilon_{j,\beta}^{(0)} + \varepsilon_{i,\alpha}^{(0)}} + \frac{1}{\varepsilon_{j,\beta}^{(0)} + \varepsilon_{i,\gamma}^{(0)}} \right]. \quad (5)$$

We remark that for $\gamma = \alpha$ we recover our previous result.

This result also includes effects of changes in neighboring-atom potentials through the first term; previously[14] we had neglected these effects. In Eq. (5) partition the Hamiltonian into two parts

$$\hat{H} = \hat{H}_i^{(\text{at})} + \hat{V}_i^{(\text{NN})}; \quad \hat{V}_i^{(\text{NN})} = \sum_{j \in \text{NN to } i} \langle \Phi_{i,\gamma} | \hat{V}_j^{(i)} | \Phi_{i,\alpha} \rangle, \quad (6)$$

where $\hat{H}_i^{(\text{at})}$ is the free-atom Hamiltonian for atom i and $\hat{V}_j^{(i)}$ is the potential at atom i due to the nearest-neighbor atom j.

(We neglect effects of more-distant neighbors.) The leading term of Eq. (5) then becomes

$$\langle \Phi_{i,\gamma} | \hat{H} | \Phi_{i,\alpha} \rangle = \varepsilon_{i,\alpha}^{(0)} \delta_{\alpha,\gamma} + \sum_{j \in \text{NN to } i} \langle \Phi_{i,\gamma} | \hat{V}_j^{(i)} | \Phi_{i,\alpha} \rangle. \quad (7)$$

In the section below we describe out treatment of nearest-neighbor potential effects.

B. Nearest-neighbor potentials

We place the ith atom at the origin and consider the potential of a nearest neighbor located at

$$\mathbf{R}_j^{(i)} = R_j^{(i)}(l_j\mathbf{e}_x + m_j\mathbf{e}_y + n_j\mathbf{e}_z), \qquad (8)$$

where (l_j, m_j, n_j) are the direction cosines from the origin to the nearest-neighbor atom j. The potential due to this neighbor is

$$\hat{V}_j^{(i)} = -\frac{e^2 Z_{eff}}{4\pi\varepsilon_0|\mathbf{r} - \mathbf{R}_j^{(i)}|}. \qquad (9)$$

In Eq. (9) we employ the permittivity of free space since the atoms are nearby and the bulk dielectric constant is a mesoscopic quantity. Screening is taken into account via the effective ionic charge, Z_{eff}; for Si $Z_{eff} \le 4$, obviously. Because the two orbitals in each matrix element in Eq. (7) are centered on the same atom, it is reasonable to assume that in a nearest-neighbor model they have decayed significantly at the neighbors. Thus in the multipole expansion of Eq. (9) we retain only the r^l/R^{l+1} terms (i.e., we assume the integral for $r: 0 \rightarrow R$ is the dominant contribution to the matrix element)

$$\frac{1}{|\mathbf{r} - \mathbf{R}|} \approx \sum_{l=0}^{\infty} \left\{ \left(\frac{4\pi}{2l+1}\right)\left(\frac{r^l}{R^{l+1}}\right) \sum_{m=-l}^{+l} Y_{l,m}(\theta,\varphi)Y_{l,m}^*(\Theta,\Phi) \right\}, \qquad (10)$$

where the angles (Θ, Φ) apply to the neighboring atom at $\mathbf{R}$. Using standard techniques[18] it is straightforward to work out the nearest-neighbor matrix elements between true atomic orbitals, leaving as fitting parameters matrix elements of the operator r^n taken between the radial parts of the two orbitals. For radial parts of orbitals $|\alpha\rangle$, $|\beta\rangle$ and radial operator, $\hat{r}$, these matrix elements are denoted

$$\overline{r_{\alpha,\beta}^n} = \langle\alpha|\hat{r}^n|\beta\rangle. \qquad (11)$$

Under the usual assumptions of empirical tight binding we expect that $0 \le \overline{r_{\alpha,\beta}^n}/R^n \le 1$. Other relationships between the parameters are likewise expected. Most obviously, based on the well-known inequality $\langle(\hat{A} - \langle\hat{A}\rangle)^2\rangle \ge 0$ we expect that $\overline{r_{d,d}^4}/R^4 \ge (\overline{r_{d,d}^2}/R^2)^2$. In addition, relationships for the p-d and d-d parameters can be deduced; these are discussed in the appendix.

In order to limit the number of parameters as much as possible we include nearest-neighbor effects only on the p and d orbitals. The results initially appear in terms of the spherical harmonics $Y_{l,m}^*(\Theta, \Phi)$, which are then rewritten in terms of the direction cosines. From these expressions one can extract effective Slater-Koster parameters[5] as discussed in the appendix. The results are

$$V_{pp\sigma}(R) = -\frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R}\left(u_0 + \frac{2}{5}\frac{\overline{r_{p,p}^2}}{R^2}\right), \qquad (12)$$

$$V_{pp\pi}(R) = -\frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R}\left(u_0 - \frac{1}{5}\frac{\overline{r_{p,p}^2}}{R^2}\right), \qquad (13)$$

$$V_{pd\sigma}(R) = -\frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R}\frac{1}{\sqrt{15}}\left(2\frac{\overline{r_{p,d}}}{R} + \frac{9}{7}\frac{\overline{r_{p,d}^3}}{R^3}\right), \qquad (14)$$

$$V_{pd\pi}(R) = -\frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R}\frac{1}{\sqrt{5}}\left(\frac{\overline{r_{p,d}}}{R} - \frac{3}{7}\frac{\overline{r_{p,d}^3}}{R^3}\right), \qquad (15)$$

$$V_{dd\sigma}(R) = -\frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R}\left(u_0 + \frac{2}{7}\frac{\overline{r_{d,d}^2}}{R^2} + \frac{2}{7}\frac{\overline{r_{d,d}^4}}{R^4}\right), \qquad (16)$$

$$V_{dd\pi}(R) = -\frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R}\left(u_0 + \frac{1}{7}\frac{\overline{r_{d,d}^2}}{R^2} - \frac{4}{21}\frac{\overline{r_{d,d}^4}}{R^4}\right), \qquad (17)$$

$$V_{dd\delta}(R) = -\frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R}\left(u_0 - \frac{2}{7}\frac{\overline{r_{d,d}^2}}{R^2} + \frac{1}{21}\frac{\overline{r_{d,d}^4}}{R^4}\right). \qquad (18)$$

In Eqs. (12)–(18) strictly speaking $u_0 = 1$; however, as this term is the same for *all* diagonal matrix elements (including s-like ones) and it vanishes for off-diagonal ones (here different orbitals p-p and d-d, and all p-d), its effect is to shift the *entire* band structure up and down without distortion and we therefore set $u_0 = 0$. The absence of this term from the strictly off-diagonal matrix elements ($V_{pd\sigma}$, $V_{pd\pi}$) is therefore no accident and indeed expected. Equations (12)–(18) are then used in the Slater-Koster[5] tables to find the matrix element between two orbitals on the same atom due to the nearest-neighbor atom located at $\mathbf{R}$. For example, to find the matrix element between a px and a py orbital due to a nearest-neighbor potential at $\mathbf{R}_j^{(i)}$ in Eq. (8) one reads off the $E_{x,y}$ entry in the Slater-Koster[5] tables

$$\langle\Phi_{i,x}|\hat{V}_j^{(i)}|\Phi_{i,y}\rangle = l_j m_j[V_{pp\sigma}(R_j^{(i)}) - V_{pp\pi}(R_j^{(i)})]. \qquad (19)$$

We remark that while these expressions take the form of two-center integrals, the signs of the matrix elements do not necessarily follow those of the usual two-center integrals because here both orbitals are centered at the origin, not one at the origin and the other at $\mathbf{R}$. We emphasize that these effective Slater-Koster[5] parameters are explicit functions of the interatomic distances $R_j^{(i)}$ and that this dependence is taken into account as discussed in Sec. III below and in the appendix. In addition, we discuss gauge invariance in the appendix, with special reference to the same-atom different-orbital parameters introduced here. As discussed in the appendix, the incompleteness of the finite basis means that the momentum and position operators in the zero-field Hamiltonian and electromagnetic coupling are different. As a result, parameters such as $\overline{r_{p,d}}$ appear only in the zero-field Hamiltonian *not* in the electromagnetic coupling terms.

C. Implementation

As in our earlier diagonal parameter strain work[14] we implement these effects as shifts to the zero-strain Hamiltonian matrix elements. We adopt this approach for the entirely practical reason that at zero-strain conventional parameter choices are often more restrictive than they need be. Two examples for diamond illustrate the point. In most tight-binding parametrizations for zero strain only one diagonal

parameter, ε_d, is used for all five d orbitals whereas two different ones are actually allowed: ε_{d1} for $\{xy, yz, zx\}$ and ε_{d2} for $\{x^2 - y^2, 3z^2 - r^2\}$. Likewise there is no requirement that the s/s^* same-atom Hamiltonian matrix element vanish, although it is nearly always set to zero. The zero-strain parameters could well alter such matrix elements and thus we implement Löwdin orbital and nearest-neighbor potential effects as shifts.

The shift to the same-atom Löwdin-basis Hamiltonian matrix element in the strained crystal (primed) is related to that in the unstrained crystal (unprimed) by

$$[\underline{H}']_{(i,\gamma),(i,\alpha)} = [\underline{H}]_{(i,\gamma),(i,\alpha)} + [\Delta\underline{H}^{(\mathrm{L})}]_{(i,\gamma),(i,\alpha)} + [\Delta\underline{V}^{(\mathrm{NN})}]_{(i,\gamma),(i,\alpha)}. \tag{20}$$

In Eq. (20) the Löwdin shift term is

$$
\begin{aligned}
[\Delta\underline{H}^{(\mathrm{L})}]_{(i,\gamma),(i,\alpha)} = &-\frac{1}{2}\sum_{j\in\mathrm{NN\ to}\ i,\beta}\left\{\left[v'_{(i,\gamma),(j,\beta)}v'_{(j,\beta),(i,\alpha)} - v_{(i,\gamma),(j,\beta)}v_{(j,\beta),(i,\alpha)}\right]\left[\frac{K_{(j,\beta),(i,\alpha)}}{\varepsilon_{j,\beta} + \varepsilon_{i,\alpha}} + \frac{K_{(i,\gamma),(j,\beta)}}{\varepsilon_{j,\beta} + \varepsilon_{i,\gamma}}\right]\right\}\\
&-\frac{1}{4}\sum_{j\in\mathrm{NN\ to}\ i,\beta}\left\{\left[v'_{(i,\gamma),(j,\beta)}v'_{(j,\beta),(i,\alpha)} - v_{(i,\gamma),(j,\beta)}v_{(j,\beta),(i,\alpha)}\right]K_{(i,\gamma),(j,\beta)}K_{(j,\beta),(i,\alpha)}\left[\frac{1}{\varepsilon_{j,\beta} + \varepsilon_{i,\alpha}} + \frac{1}{\varepsilon_{j,\beta} + \varepsilon_{i,\gamma}}\right]\right\},
\end{aligned}
\tag{21}
$$

where in Eq. (21) primes denote nearest-neighbor Hamiltonian matrix elements in the strained crystal, and as in Ref. 14 in the denominators we replace the atomic energies by their vacuum-referenced Löwdin orbital counterparts in the unstrained crystal: $\varepsilon_{i,\alpha}^{(0)} \rightarrow \varepsilon_{i,\alpha}$. The shift due to nearest-neighbor potentials in Eq. (20) is, from Eq. (7),

$$
\begin{aligned}
[\Delta\underline{V}^{(\mathrm{NN})}]_{(i,\gamma),(i,\alpha)} = \sum_{j\in\mathrm{NN\ to}\ i}&[\langle\Phi_{i,\gamma}|\hat{V}_j'^{(i)}|\Phi_{i,\alpha}\rangle\\
&- \langle\Phi_{i,\gamma}|\hat{V}_j^{(i)}|\Phi_{i,\alpha}\rangle],
\end{aligned}
\tag{22}
$$

where once again primes denote the potentials in the strained crystal. As described above, the atomic-orbital basis matrix elements in Eq. (22) may be calculated from the Slater-Koster[5] tables and the effective parameters of Eqs. (12)–(18).

III. RESULTS

A. Bulk crystals

The optimized strain parameters for Si are listed in Table I (strain exponents), Table II (shift constants), and Table III

(nearest-neighbor potential radii and effective charge); the optimization was carried out using our genetic algorithm (GA) (Refs. 9 and 19) optimization framework, described in detail in Ref. 9. The GA mimics evolutionary processes in biological systems by varying parameters through mutation and crossover operations. Initial populations of the parameters being optimized are used to calculate trial observables (band-edge energies and effective masses) under varying strain conditions, and those which are best fits are allowed to cross and mutate (i.e., reproduce). The quality of the fit is determined by a weighted average of the variances between the actual and targeted values. The reproduction-and-culling cycle continues for a selected number of generations and the best set found is returned. The optimization process is computationally extremely intensive, with millions of trial candidates being evaluated on parallel computer clusters. The optimization process is performed through the statistical evaluation using a GA since there is no general procedure for generating the Hamiltonian matrix elements of a tight-binding model with a given basis set and extent of near-neighbor interactions directly from *ab initio* wave functions and Hamiltonians.

TABLE I. Scaling exponents (dimensionless) for Si two-center integrals from Ref. 23.

Exponent, η		Exponent, η	
$ss\sigma$	0.56247	$pp\sigma$	0.20000
$s^*s^*\sigma$	0.19237	$pp\pi$	1.67770
$ss^*\sigma$	0.13203	$pd\sigma$	0.20000
$sp\sigma$	2.36548	$pd\pi$	4.43250
$s^*p\sigma$	0.34492	$dd\sigma$	0.10000
$sd\sigma$	2.56720	$dd\pi$	6.00000
$s^*d\sigma$	1.08601	$dd\delta$	5.99970

TABLE II. Löwdin shift constants (dimensionless) for Si. The vacuum-referenced Löwdin orbital energies are found by applying a downward shift $E_{shift}=27.0$ eV to the zero-strain diagonal (i.e., same-atom, same-orbital) parameters. These constants are optimized for use with the diagonal parameters and two-center integrals of Ref. 23.

Shift const., C		Shift const., C	
ss	1.6788	sd	0.0000
s^*s^*	0.7777	s^*d	0.3421
ss^*	1.7843	pp	0.0000
sp	0.4801	pd	0.0580
s^*p	3.5888	dd	4.3269

TABLE III. Dimensionless constants for the nearest-neighbor potential shift terms in Si; $R_0 = \sqrt{3}a_0/4$ is the zero-strain nearest-neighbor distance and $a_0 = 0.543$ nm is the conventional unit-cell cube edge. These parameters are optimized for use with the unstrained Si parameters of Ref. 23.

Parameter		Parameter	
$\overline{r_{p,p}^2}/R_0^2$	0.1227	$\overline{r_{d,d}^2}/R_0^2$	0.5147
$\overline{r_{p,d}}/R_0$	0.1249	$\overline{r_{d,d}^4}/R_0^4$	0.7661
$\overline{r_{p,d}^3}/R_0^3$	0.0000	Z_{eff}	3.0000

The targets for and conditions under which the optimization was carried out are as follows. The band-edge energies were fitted to van de Walle's model solid theory,[20] while effective masses were fitted to VASP (Ref. 21) calculations. As before, spin orbit is treated using Chadi's[22] approach, while the diagonal parameters and two-center integrals are given in Ref. 23. In the case of shear strains, displacements of the two atoms within a unit cell are calculated using Kleinman's model,[24] with internal displacement parameter $\zeta_{Si} = 0.53$.

A few comments on the parameters in Tables II and III are in order. The shift constants in Table II are defined identically to those in our earlier work;[14] the only difference is that now we apply them to both off-diagonal and diagonal same-atom matrix elements. Thus the constants, C, in Table II are related to the overlap constants, K, as before: K is the positive root of $K^2 + 2K - 2C = 0$. In using Table III it must be kept in mind that its parameters (except of course Z_{eff}) are scaled to the unstrained Si nearest-neighbor distance, R_0. Thus they must be appropriately scaled when used in the expressions for the effective Slater-Koster[5] parameters, Eqs. (12)–(18), as must the common $1/R$ prefactor. This procedure is illustrated in the appendix.

Figure 1 shows the uniaxial [001] strain behavior of Si band edges with the parameters of Tables I–III, i.e., $\varepsilon_{xx} = \varepsilon_{yy} = -\varepsilon_{zz}c_{11}/(2c_{12})$. Solid lines are the present tight-binding model and solid symbols are the model solid theory of Ref. 20. The conduction-band minima and valence-band
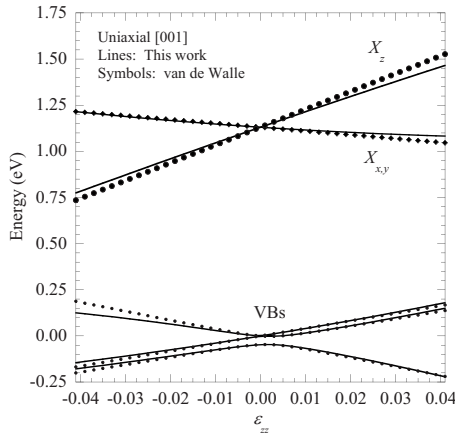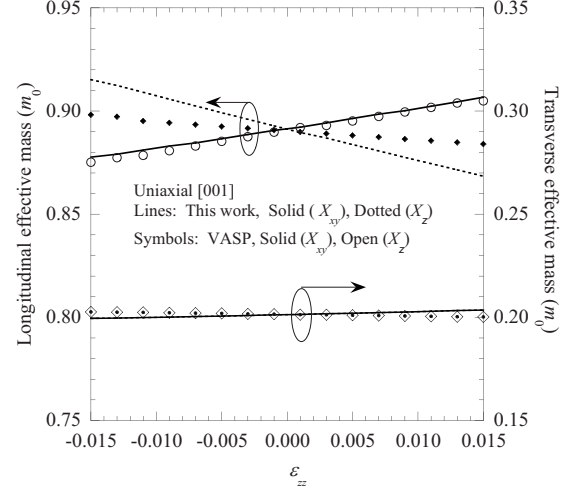

FIG. 2. Conduction-band masses of Si under uniaxial [001] strain as reproduced by this model ($X_{x,y}$: solid lines and X_z: dotted lines) and as predicted by VASP (Ref. 21) ($X_{x,y}$: solid symbols and X_z: open symbols), in units of the free-electron mass, m_0. Although the longitudinal masses in the two calculations follow opposite trends, the maximum difference between the tight-binding and VASP (Ref. 16) results is less than 4.5%.

maxima are well reproduced over the range of strain. Figure 2 presents the longitudinal and transverse conduction-band effective masses. Lines are the masses reproduced by this model, symbols are VASP (Ref. 21) results. Solid lines and symbols refer to the $X_{x,y}$ valleys while dotted lines and open symbols refer to the X_z valleys. The transverse masses (more critical for transport) reproduce the VASP (Ref. 21) results very well. There is a small discrepancy in the longitudinal masses, with the two valleys in tight binding following opposite trends from their VASP (Ref. 21) counterparts. The disagreement is at most 4.5%, however, and in any event the longitudinal masses are much less relevant for transport. Figure 3 shows the uniaxial [110] strain behavior of the band edges; the experimentally accessible region is indicated by

FIG. 3. Conduction- and valence-band edges of Si under uniaxial [110] strain as reproduced by this model (lines) and as predicted by the 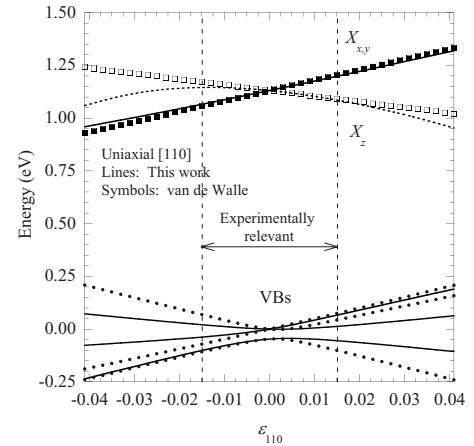model solid theory of Ref. 20 (symbols). The experimentally relevant region, $|\varepsilon_{110}| \lesssim 0.015$, is indicated by vertical dashed lines.

FIG. 1. Conduction- and valence-band edges of Si under uniaxial [001] strain as reproduced by this model (lines) and as predicted by the model solid theory of Ref. 20 (symbols).
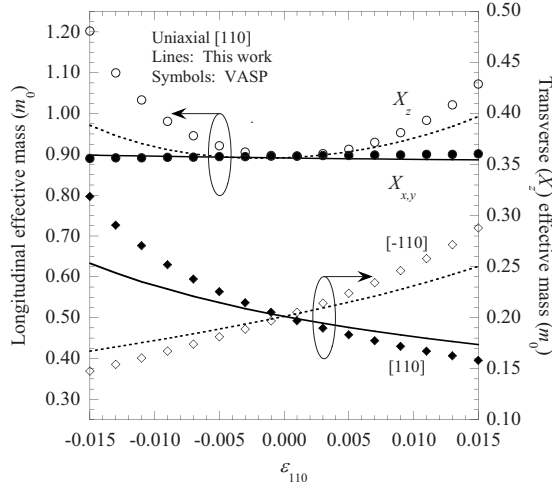
FIG. 4. Conduction-band masses of Si under uniaxial [110] strain as reproduced by this model (lines) and as predicted by VASP (Ref. 21) (symbols), in units of the free-electron mass, m_0. The transverse masses are those of the X_z valleys.

dotted vertical lines. (Experimental strains beyond this range risk damage to the sample.) The conduction bands are generally better reproduced than are the valence bands. This result is a compromise due to our more heavily weighting the transverse mass splitting than the band-edge behavior in the optimization. Figure 4 shows the conduction-band effective-mass behavior over the range of strain relevant for experiments. The transverse mass splitting of the X_z valleys is well reproduced, while the longitudinal mass splitting is underestimated at higher strains. Again, however, it is the transverse mass behavior which is critical for transport calculations, and it was most heavily weighted in our optimization. Figures 5 and 6 show the conduction- and valence-band edges and conduction-band effective masses, respectively, under hydrostatic strain. In both cases a very good fit is achieved.

B. Device calculations

We next consider the behavior of a Si, double gate, and ultrathin-body field effect transistor. The physical gate length
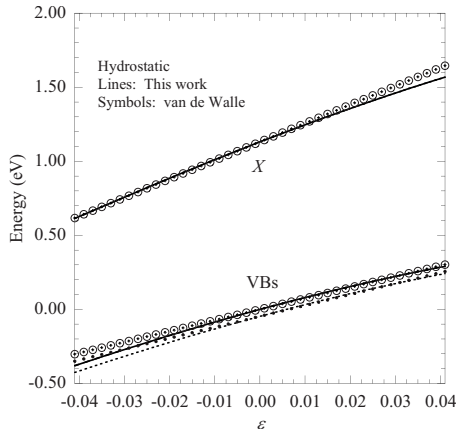

FIG. 5. Conduction- and valence-band edges of Si under hydrostatic strain as reproduced by this model (lines) and as predicted by the model solid theory of Ref. 20 (symbols).
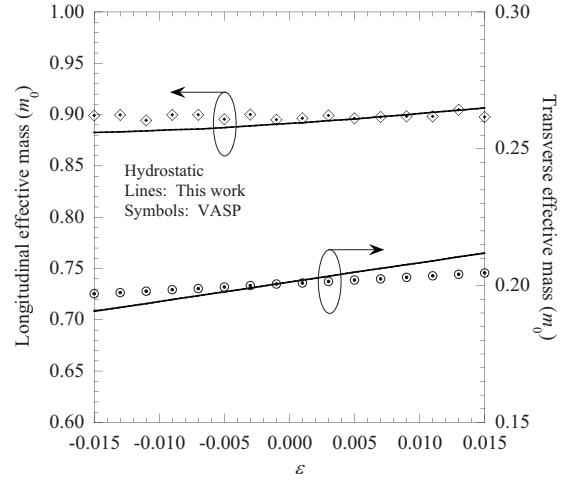

FIG. 6. Conduction-band masses of Si under hydrostatic strain as reproduced by this model (lines) and as predicted by VASP (Ref. 21) (symbols) in units of the free-electron mass, m_0.

is 22 nm with source and drain N-type doped to 10^{20} cm^{-3}. The channel and metallic gate are separated by SiO$_2$ oxide ($\varepsilon_{rel}=3.9$) with a thickness of 1.3 nm. Transport occurs along [110], quantization along [001], and the width is along [$\bar{1}$10]. The operating conditions are $V_{gs}=V_{ds}=V_{DD}=1$ V at which we calculate the ON current, I_d, under various strain conditions. We also calculate the change from the zero-strain current, I_{d0}, defined as $(I_d-I_{d0})/I_{d0}$ and expressed as a percentage. As a reference the OFF current (I_d at $V_{ds}=V_{DD}$ and $V_{gs}=0$ V) of all devices is set to the same value before extracting the ON current. The drain currents (open symbols) and percent changes (closed symbols) are plotted in Fig. 7; lines are guides to the eyes. The saturation of the current under biaxial [110] strain is similar to that in the calculations of Ref. 10, although in our calculations the biaxial and
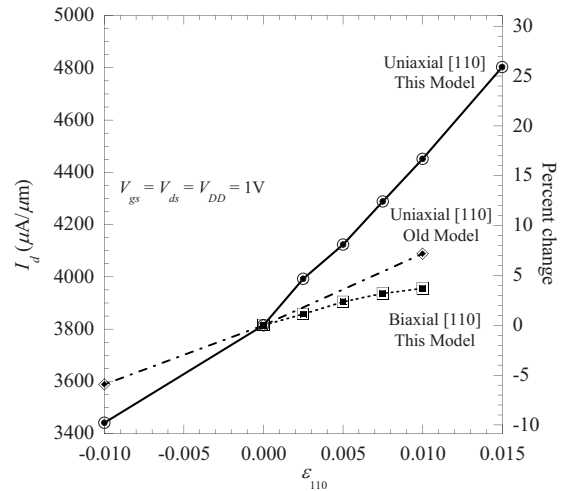

FIG. 7. Drain ON current, I_d (open symbols) as a function of strain and percent change relative to the drain ON current at zero strain (closed symbols) for a Si, double-gate, ultrathin-body MOSFET (see text); lines are guides to the eyes. The operating conditions are: $V_{gs}=V_{ds}=V_{DD}=1$ V. The curve labeled "Old Model" is a calculation using the strain model of Ref. 14.

uniaxial curves do not cross. Under uniaxial [110] strain this model predicts a much larger change in current than does our previous model. This large change in current (about 17% at 1% strain) is less than seen in the device calculations of Uchida *et al.*,[10] but is nevertheless comparable to experimental measurements.[11,25] In comparing our results to those of Ref. [10], it is vital to keep in mind that our calculation is multiband while theirs is effective mass and that we include all valleys (even those highly unfavorable for transport) while it is not clear that they include all six valleys. The more relevant comparison of the present model is therefore to experiments, and in this regard it is much improved and gives believable results.

IV. CONCLUSIONS

We have presented a tight-binding strain model for nearest-neighbor bases which explicitly includes effects of both Löwdin orbital changes and displacement of nearest-neighbor potentials. The model allows for same-atom, different-orbital Hamiltonian matrix elements and provides a much better treatment of the technologically important case of uniaxial [110] strain. Unlike other models which include these matrix elements,[15] in this model it is easy to deduce reasonable limits on the parameters. We have illustrated the method with a strain parametrization for Si and have employed the resulting parameters to calculate the drain current of a MOSFET under both uniaxial [110] and biaxial [110] strain. This model predicts a much larger change in the drain current than does out earlier strain model,[14] and this larger change is in reasonable agreement with experiments.[25] This model should therefore be useful for multiband Si device calculations.

APPENDIX

1. Nearest-neighbor potential effective Slater-Koster parameters

The nearest-neighbor potential parameters can be converted from expressions in terms of spherical harmonics into the form of Slater-Koster parameters by writing the spherical harmonics in terms of direction cosines and then, if necessary, using the direction cosine identity $l^2 + m^2 + n^2 = 1$. For

the p-p matrix elements the procedure is easily illustrated, while for some of the d-d matrix elements it can be lengthy. (For the d-d matrix elements the Slater-Koster[5] tables tend to be more complicated than what results from simply expressing spherical harmonics in terms of direction cosines.) As an example, consider the x-x matrix element on the ith atom due to the nearest neighbor, j, at $(R_j^{(i)}, \Theta_j^{(i)}, \Phi_j^{(i)})$

$$
\begin{aligned}
\langle \Phi_{i,x} | \hat{V}_j^{(i)} | \Phi_{i,x} \rangle = &- \frac{e^2 Z_{eff}}{4\pi\varepsilon_0} \frac{2\sqrt{\pi}}{R_j^{(i)}} Y_{0,0}^*(\Theta_j^{(i)}, \Phi_j^{(i)}) \\
&- \frac{e^2 Z_{eff}}{4\pi\varepsilon_0} \frac{\overline{r_{p,p}^2}}{(R_j^{(i)})^3} \left\{ -\frac{2\sqrt{5\pi}}{25} Y_{2,0}^*(\Theta_j^{(i)}, \Phi_j^{(i)}) \right. \\
&\left. + \frac{\sqrt{30\pi}}{25} [Y_{2,2}^*(\Theta_j^{(i)}, \Phi_j^{(i)}) + Y_{2,-2}^*(\Theta_j^{(i)}, \Phi_j^{(i)})] \right\}.
\end{aligned} \tag{A1}
$$

This matrix element is first simplified by expressing the spherical harmonics in terms of direction cosines (l_j, m_j, n_j) instead of the angles. After a little simplification one obtains

$$
\begin{aligned}
\langle \Phi_{i,x} | \hat{V}_j^{(i)} | \Phi_{i,x} \rangle = &- \frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R_j^{(i)}} \left[1 + \frac{1}{10}(1 - 3n_j^2 + 3l_j^2 \right. \\
&\left. - 3m_j^2) \frac{\overline{r_{p,p}^2}}{(R_j^{(i)})^2} \right].
\end{aligned} \tag{A2}
$$

Use of the identity for direction cosines leads to

$$
\langle \Phi_{i,x} | \hat{V}_j^{(i)} | \Phi_{i,x} \rangle = - \frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R_j^{(i)}} \left[\left(1 - \frac{1}{5} \frac{\overline{r_{p,p}^2}}{(R_j^{(i)})^2} \right) + \frac{3}{5} l_j^2 \frac{\overline{r_{p,p}^2}}{(R_j^{(i)})^2} \right]. \tag{A3}
$$

On the other hand, the Slater-Koster[5] tables give

$$
E_{x,x} = V_{pp\pi} + l^2(V_{pp\sigma} - V_{pp\pi}). \tag{A4}
$$

Because the angular dependence must be identical in both cases we can recover the effective Slater-Koster parameters by equating coefficients of like angular expressions. From the angle-independent term we obtain immediately

$$
V_{pp\pi}(R_j^{(i)}) = - \frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R_j^{(i)}} \left(1 - \frac{1}{5} \frac{\overline{r_{p,p}^2}}{(R_j^{(i)})^2} \right). \tag{A5}
$$

Using Eq. (A5) together with Eqs. (A3) and (A4) in the l^2 term then gives

$$
V_{pp\sigma}(R_j^{(i)}) = - \frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R_j^{(i)}} \left(1 + \frac{2}{5} \frac{\overline{r_{p,p}^2}}{(R_j^{(i)})^2} \right). \tag{A6}
$$

The expressions for other matrix elements can then be checked with these definitions.

In using the parameters of Table III with these expressions, we emphasize that the distances must be scaled as the parameters are normalized to the unstrained Si nearest-neighbor distance. For example, one computes Eq. (A6) using the parameters of Table III as
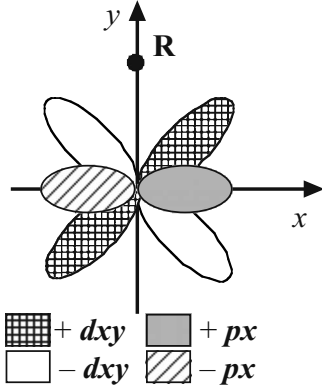
FIG. 8. Schematic of the same-atom matrix element $\langle \Phi_{i,x} | \hat{V}_j^{(i)} | \Phi_{i,xy} \rangle$ for $l_j = n_j = 0$, $m_j = 1$; **R** is the location of the nearest-neighbor potential. Because the potential is attractive and nearest to a region of positive orbital product ($+/+$, $-/-$ orbital overlap), this matrix element is expected to be negative.

$$V_{pp\sigma}(R_j^{(i)}) = -\frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R_0} \left(\frac{R_0}{R_j^{(i)}} \right) \left[1 + \frac{2}{5} \left(\frac{\overline{r_{p,p}^2}}{R_0^2} \right) \left(\frac{R_0}{R_j^{(i)}} \right)^2 \right]. \tag{A7}$$

The other effective Slater-Koster parameters are scaled in a similar manner.

Finally, because the nearest-neighbor potentials are taken between atomic states, deducing expected signs of matrix elements, and from these relationships among the parameters is reasonably straightforward. Because of the approximations made, such as retaining only the r^l / R^{l+1} terms in the multipole expansion, these guides cannot be interpreted as absolutes. They are nevertheless very useful in the sense that nearest-neighbor potential parameters, which are far out of compliance with them are likely unphysical.

Figure 8 illustrates the same-atom parameter $\langle \Phi_{i,x} | \hat{V}_j^{(i)} | \Phi_{i,xy} \rangle$ for $l_j = n_j = 0$, $m_j = 1$; the nearest-neighbor potential is located at point **R**. Because the potential is attractive and nearest to regions of positive orbital product ($+/+$, $-/-$ orbital overlap), the entire matrix element is expected to be negative. In fact, Fig. 8 shows that for any $m_j > 0$ the matrix element should be negative (semidefinite). From the Slater-Koster[5] tables

$$E_{x,xy} = m V_{pd\pi} + m l^2 (\sqrt{3} V_{pd\sigma} - 2 V_{pd\pi}). \tag{A8}$$

Equations (14) and (15) give

$$\sqrt{3} V_{pd\sigma} - 2 V_{pd\pi} = -\frac{e^2 Z_{eff}}{4\pi\varepsilon_0 R} \frac{3\sqrt{5}}{7} \frac{\overline{r_{pd}^3}}{R^3}. \tag{A9}$$

The second term of Eq. (A8) is therefore negative (semidefinite) so long as $\overline{r_{pd}^3} \geq 0$. The first term is negative (semidefinite) so long as $V_{pd\pi} \leq 0$, which, according to Eq. (15), leads to the conclusion

$$\frac{\overline{r_{p,d}}}{R} \geq \frac{3}{7} \frac{\overline{r_{p,d}^3}}{R^3}. \tag{A10}$$

A similar construction for the *d-d* matrix element $E_{xy,yz}$ (not shown) leads to the inequality

$$\frac{\overline{r_{d,d}^2}}{R^2} \geq \frac{5}{9} \frac{\overline{r_{d,d}^4}}{R^4}. \tag{A11}$$

The parameters in Table III satisfy these relations. Again, we emphasize that due to the approximations made these relations cannot be taken as absolutes. However, they do indicate that parameter sets which are far from satisfying them likely have some unphysical parameters.

2. Gauge invariance

The issue of gauge invariance in empirical tight-binding models is a subtle one and has been much discussed in the literature.[26–30] The key to understanding the issue is the recognition that different position and momentum operators appear in the zero-field Hamiltonian and in the electromagnetic coupling operators (vector potential and momentum). Drawing such a distinction is necessary because the finite tight-binding basis is incomplete and cannot represent spatial differentiation, only discrete atomic separations in operators external to the Hamiltonian. This distinction between the different treatments of the position and momentum operators in the zero-field Hamiltonian and the electromagnetic coupling terms is regularly made when one adopts Chadi's[22] spin-orbit treatment. As we show below, so long as one adopts the minimal prescription for the position operator in calculating external quantities (electromagnetic fields, currents, etc.) gauge invariance is unaffected by strain-induced same-atom, different-orbital Hamiltonian matrix elements, such as the $\overline{r_{p,d}}$ parameter above.

As discussed in Refs. 26–29, the only prescription for the position operator for *external* fields compatible with gauge invariance in an arbitrary orthogonal basis is one which gives the position of the atom on which the orbital is centered. For the Löwdin orbital basis, this prescription is

$$\langle \varphi_{i,\alpha} | \hat{r}^{(\gamma)} | \varphi_{j,\beta} \rangle = R_i^{(\gamma)} \delta_{i,j} \delta_{\alpha,\beta}, \quad \gamma \in \{x, y, z\}. \tag{A12}$$

(Reference 29 shows that same-atom, different-orbital Löwdin-basis position matrix elements can be compatible with gauge invariance, but only if the choice of basis states is restricted to certain orbitals.) Reference 30 shows that vanishing same-atom, different-orbital Löwdin-basis position matrix elements can be compatible with *finite* same-atom, different-orbital atomic-basis position matrix elements. We emphasize as well that the same-atom different-orbital position matrix elements such as $\overline{r_{p,d}}$ are taken between *atomic* orbitals not the orthogonal Löwdin-basis orbitals.

Because the only new interactions introduced are same-atom, different-orbital, we need only check same-atom electromagnetic coupling matrix elements to establish gauge invariance. With definition in Eq. (A12) for the position operator for *external* interactions and the power-series representation for the electromagnetic coupling in Ref. 27, the leading term (disregarding constants) is

$$\langle\varphi_{i,\alpha}|\hat{A}^{(\gamma)}\hat{p}^{(\gamma)}+\hat{p}^{(\gamma)}\hat{A}^{(\gamma)}|\varphi_{i,\beta}\rangle = 2\hat{A}^{(\gamma)}(\mathbf{R}_i,t)\langle\varphi_{i,\alpha}|\hat{p}^{(\gamma)}|\varphi_{i,\beta}\rangle. \tag{A13}$$

The momentum matrix element in Eq. (A13) vanishes, however,

$$\langle\varphi_{i,\alpha}|\hat{p}^{(\gamma)}|\varphi_{i,\beta}\rangle = \frac{m_0}{i\hbar}\langle\varphi_{i,\alpha}|[\hat{r}^{(\gamma)},\hat{H}_0]|\varphi_{i,\beta}\rangle = \frac{m_0}{i\hbar}(R_i^{(\gamma)}-R_i^{(\gamma)})$$

$$\times\langle\varphi_{i,\alpha}|\hat{H}_0|\varphi_{i,\beta}\rangle = 0, \tag{A14}$$

where $\hat{H}_0$ is the Hamiltonian with $\mathbf{A}(\mathbf{r},t)=\mathbf{0}$ for the crystal (including any strain). Additional electromagnetic coupling terms involve more commutators with the position operators and therefore likewise vanish for orbitals on the same atom. Thus the nonvanishing same-atom Hamiltonian matrix elements do not directly affect the optical localized-orbital matrix elements. (They do indirectly affect optical properties by altering the zero-field initial and final states.) The localized-orbital optical matrix elements are of course directly affected by the displacement of the neighboring atoms and scaling of neighboring-atom Hamiltonian matrix elements.

It is easily demonstrated that the matrix elements of a scalar potential are likewise unaffected by the same-atom, different-orbital Hamiltonian matrix elements. Again adopting the gauge-invariant prescription for the external potential position operator, Eq. (A12), the matrix elements of an external scalar potential, $V_{\text{ext}}(\mathbf{r},t)$, in the Löwdin basis are (again disregarding constants)

$$\langle\varphi_{i,\alpha}|\hat{V}_{\text{ext}}(\mathbf{r},t)|\varphi_{j,\beta}\rangle = V_{\text{ext}}(\mathbf{R}_j,t)\langle\varphi_{i,\alpha}|\varphi_{j,\beta}\rangle = V_{\text{ext}}(\mathbf{R}_j,t)\delta_{i,j}\delta_{\alpha,\beta}, \tag{A15}$$

where as usual one treats $\hat{V}_{\text{ext}}(\mathbf{r},t)$ as a power series in $(\hat{\mathbf{r}},t)$. As Eq. (A15) demonstrates, the matrix elements of an external scalar potential are identical to those in the gauge-invariant treatment of Refs. 26–28.

Finally, we emphasize that the distinction between position and momentum operators *incorporated into* the zero-field Hamiltonian and those *external* to it is not radical or unusual. It is customarily made in adopting Chadi's[22] treatment of the spin-orbit interaction. Likewise, it is difficult to see how the position and momentum operators in Eqs. (A12)–(A15) could lead to the matrix elements of $\hat{H}_0 = \hat{\mathbf{p}}^2/(2m_0)+\hat{V}(\mathbf{r})$ with or without strains. Hence gauge invariance is maintained so long as one adopts the minimal prescription for the position operator, Eq. (A12) for *external* interactions.

---

[1] R. C. Bowen, G. Klimeck, R. Lake, W. R. Frensley, and T. Moise, J. Appl. Phys. **81**, 3207 (1997).

[2] M. Usman, H. Ryu, I. Woo, D. S. Ebert, and G. Klimeck, IEEE Trans. Nanotechnol. **8**, 330 (2009).

[3] N. Kharche, M. Prada, T. B. Boykin, and G. Klimeck, Appl. Phys. Lett. **90**, 092109 (2007).

[4] G. P. Lansbergen, R. Rahman, C. J. Wellard, P. E. Rutten, J. Caro, N. Collaert, S. Biesmans, I. Woo, G. Klimeck, L. C. L. Hollenberg, and S. Rogge, Nat. Phys. **4**, 656 (2008).

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

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

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

[8] T. B. Boykin, J. P. A. van der Wagt, and J. S. Harris, Jr., Phys. Rev. B **43**, 4777 (1991); T. B. Boykin, *ibid.* **54**, 8107 (1996).

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

[10] K. Uchida, T. Krishnamohan, K. C. Saraswat, and Y. Nishi, Tech. Dig. - Int. Electron Devices Meet. 2005, 129.

[11] S. E. Thompson, S. Suthram, Y. Sun, G. Sun, S. Parthasarathy, M. Chu, and T. Nishida, Tech. Dig. - Int. Electron Devices Meet. 2006, 1.

[12] Walter A. Harrison, *Elementary Electronic Structure* (World Scientific, New Jersey, 1999).

[13] C. Tserbak, H. M. Polatoglou, and G. Theodorou, Phys. Rev. B **47**, 7104 (1993).

[14] T. B. Boykin, G. Klimeck, R. C. Bowen, and F. Oyafuso, Phys. Rev. B **66**, 125207 (2002).

[15] Y. M. Niquet, D. Rideau, C. Tavernier, H. Jaouen, and X. Blasé, Phys. Rev. B **79**, 245201 (2009).

[16] P.-O. Löwdin, J. Chem. Phys. **18**, 365 (1950).

[17] J. M. Wills and W. A. Harrison, Phys. Rev. B **28**, 4363 (1983).

[18] M. Weissbluth, *Atoms and Molecules* (Academic, San Diego, CA, 1978).

[19] G. Klimeck, R. C. Bowen, T. B. Boykin, and T. A. Cwick, Superlattices Microstruct. **27**, 519 (2000).

[20] C. G. Van de Walle, Phys. Rev. B **39**, 1871 (1989).

[21] VASP is documented in: G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993); **49**, 14251 (1994).

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

[23] T. B. Boykin, G. Klimeck, and F. Oyafuso, Phys. Rev. B **69**, 115201 (2004).

[24] L. Kleinman, Phys. Rev. **128**, 2614 (1962).

[25] S. E. Thompson (private communication).

[26] M. Graf and P. Vogl, Phys. Rev. B **51**, 4940 (1995).

[27] T. B. Boykin, R. C. Bowen, and G. Klimeck, Phys. Rev. B **63**, 245314 (2001).

[28] T. B. Boykin and P. Vogl, Phys. Rev. B **65**, 035202 (2001).

[29] B. A. Foreman, Phys. Rev. B **66**, 165212 (2002).

[30] T. Sandu, Phys. Rev. B **72**, 125105 (2005).