PHYSICAL REVIEW B VOLUME 62, NUMBER 8 15 AUGUST 2000-II

Method for tight-binding parametrization: Application to silicon nanostructures

Y. M. Niquet,* C. Delerue, G. Allan, and M. Lannoo
Institut d'Electronique et de Microélectronique du Nord, Département ISEN, Boîte Postale 69, F-59652 Villeneuve d'Ascq Cedex, France

(Received 18 February 2000)

We propose a method for tight-binding parametrization, designed to give accurate results in the calculation of confined edge states in semiconductor nanostructures of any size. Indeed, this improved tight-binding description accurately reproduces the bulk effective masses as well as the overall band structure. We apply it to the specific case of silicon. The electronic states of silicon nanostructures (films, wires, and dots), with various shapes and orientations, are calculated over large range of sizes (1–12 nm), including spin orbit. Accurate analytical laws for the confinement energies, valid over the whole range of sizes, are derived. Consistent comparison with the effective mass and $\mathbf{k}\cdot\mathbf{p}$ methods show that these are only of semiquantitative value even for sizes as large as 8 nm. The reasons for the failure of these techniques is analyzed in detail.

I. INTRODUCTION

Recent developments in the field of Si nanostructures[1–3] have made possible devices with feature sizes below 10 nm. These devices have shown exciting low-temperature transport properties[4] with promising applications in microelectronics (single-electron transistors and memories). The most appealing challenge is now to achieve reliable room-temperature operation. In this context, simulation is one of the keys to a better understanding of the underlying physics and optimization of these devices. An accurate and efficient description of the electronic properties of Si nanostructures with arbitrary geometries in the range 1–10 nm is thus needed. However, such a description is still missing.

Indeed, *ab initio*, self-consistent methods such as the local-density approximation (LDA) can only be used for small clusters (<1000 atoms) with high symmetry.[5] They are not suitable for the computation of transport properties in realistic situations. Semiempirical, non self-consistent methods [such as pseudopotential[6,7] (PP's), tight-binding[8–11] (TB), or $\mathbf{k}\cdot\mathbf{p}$ (Ref. 12)] can solve much larger problems. Semiempirical methods are designed to make the best possible approximation to the self-consistent one-particle Hamiltonian H_0 in bulk material, either in the whole first Brillouin zone (PP, TB) or around specific k points ($\mathbf{k}\cdot\mathbf{p}$). They involve adjustable parameters (e.g., effective masses, TB interaction parameters etc.) that are fitted to experimental data or *ab initio* bulk band structures. These parameters are then transferred to the nanostructures (i.e., $H=H_0$ inside the nanostructure) with appropriate boundary conditions. The better the bulk description and boundary conditions, the better the electronic structure we expect in nanostructures.

Our aim in this work is to present an improved TB description designed to give accurate results over the whole range of sizes. For this, we fit the TB parameters not only on bulk band energies in the first Brillouin zone but also on effective masses. This was not the case of previous TB treatments, but is essential if one wants to obtain accurate results for large size nanostructures where the effective-mass approximation (EMA) or $\mathbf{k}\cdot\mathbf{p}$ models become exact. We also include spin-orbit coupling. From improved computer codes, we are able to apply the TB treatment to nanostructures with feature size in the range 1–12 nm, which to our knowledge has never been achieved previously. This extended range allows us to study precisely the overlap region where both TB and $\mathbf{k}\cdot\mathbf{p}$ or EMA methods are accurate. We provide analytic fits to the confinement energies which are practically exact over the whole range of sizes. We finally give the structure of low-lying excited states and compare with $\mathbf{k}\cdot\mathbf{p}$ and EMA methods.

We start by giving details about the TB parametrization, and compare the full results with $\mathbf{k}\cdot\mathbf{p}$ or EMA methods near the band extrema. We then consider various nanostructures establishing analytical laws for the confinement energies. We finally discuss the origins of the failure of $\mathbf{k}\cdot\mathbf{p}$ and EMA methods, as well as the critical size at which this occurs.

II. sp^3 TIGHT-BINDING MODEL

A. Tight-binding interaction parameters

TB parameters are usually fitted on bulk band structures calculated with *ab initio* methods (e.g. LDA, corrected of the band-gap problem), or even with other semiempirical methods like pseudopotentials.[13–16] TB models can provide a rather good description of the valence bands (VB's) and lowest conduction bands (CB's), with a rms energy error distributed over the whole first Brillouin zone. However, no special attention is usually paid to the description of the neighborhood of the valence-band maximum (VBM) at Γ point in Si and conduction-band minima (CBM) near X point in Si. These points are of prime importance in calculation of confined edge states, at least in large nanostructures. Indeed, most TB models do not reproduce the high anisotropy of both CB and VB effective masses. Although these models did prove to give satisfactory results in highly confined systems, they clearly do not extrapolate to $\mathbf{k}\cdot\mathbf{p}$ in large nanostructures. As an example, the orthogonal third-nearest-neighbor sp^3 TB model of Ref. 15 gives $\gamma_2 = 1.233$ and $m_l^* = 0.567$, to be compared with the experimental values[17] $\gamma_2 = 0.320$ and $m_l^* = 0.916$.

Our way to cure this problem is to fit TB parameters on bulk band energies as well as CB/VB effective masses. The total rms error is thus a weighted average of the rms error on

TABLE I. Third-nearest-neighbor (NN) TB parameters for silicon and first NN TB parameters for Si-H. The notation is that of Slater and Koster, Si-H parameters being given in terms of two-center integrals. Neighbor positions are given in units of $a/4$. Δ is the spin-orbit coupling parameter.

Si $3^{\rm rd}$ NN TB parameters			
$E_{ss}[000]$	-6.17334 eV	$E_{ss}[111]$	-1.78516 eV
$E_{pp}[000]$	2.39585 eV	$E_{sx}[111]$	0.78088 eV
		$E_{xx}[111]$	0.35657 eV
Δ	0.04500 eV	$E_{xy}[111]$	1.47649 eV
$E_{ss}[220]$	0.23010 eV	$E_{ss}[311]$	-0.06857 eV
$E_{sx}[220]$	-0.21608 eV	$E_{sx}[311]$	0.25209 eV
$E_{sx}[022]$	-0.02496 eV	$E_{sx}[113]$	-0.17098 eV
$E_{xx}[220]$	0.02286 eV	$E_{xx}[311]$	0.13968 eV
$E_{xx}[022]$	-0.24379 eV	$E_{xx}[113]$	-0.04580 eV
$E_{xy}[220]$	-0.05462 eV	$E_{xy}[311]$	-0.03625 eV
$E_{xy}[022]$	-0.12754 eV	$E_{xy}[113]$	0.06921 eV
Si-H $1^{\rm st}$ NN TB parameters:			
E_H	0.17538 eV	$V_{ss\sigma}$	-4.12855 eV
		$V_{sp\sigma}$	3.72296 eV

bulk band energies and on effective masses. It is minimized with a conjugate gradient algorithm. Bulk band energies are selected from a GW band structure,[18,19] which is the best available at the moment. An orthogonal sp^3 TB model with up to third-nearest-neighbor interactions and three-center integrals[15] is considered here. The set of 20 TB parameters is reported in Table I. TB bulk dispersion relations are plotted in Fig. 1; TB effective masses and Luttinger parameters are reported in Table II. As can be seen from Fig. 1 and Table II, the overall quality of the fit is excellent. In particular, the valence band anisotropy is well reproduced, the ratio $(\gamma_3-\gamma_2)/\gamma_1$ being close to the experimental value. We may therefore expect better hole wave functions in nanostructures. This is, to our knowledge, the best description of the VBM and CBM obtained so far in Si with a sp^3 TB model. This TB band structure is close to the one of Ref. 15 in the rest of the first Brillouin zone.

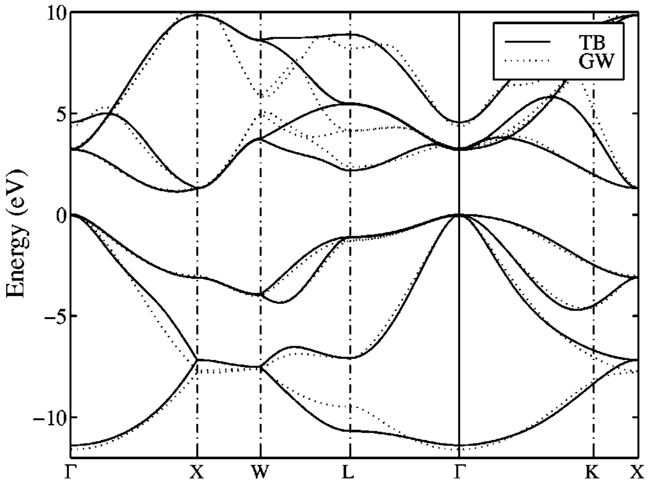

FIG. 1. Band structure of bulk Si in the orthogonal third-nearest-neighbor TB and GW models. TB parameters are given in Table I.

TABLE II. Experimental and TB band-gap energy, effective masses, and valence-band Luttinger parameters. A lattice parameter $a=5.431$ Å was assumed in the calculation of TB effective masses.

	Experiment (Ref. 17)	TB model
	Indirect band-gap energy (at $0.832\ \Gamma X$):	
E_g	1.170 eV	1.143 eV
	Conduction-band effective masses:	
m_l^*	0.916 m_0	0.918 m_0
m_t^*	0.191 m_0	0.191 m_0
	Valence-band Luttinger parameters:	
γ_1	4.270	4.271
γ_2	0.320	0.408
γ_3	1.458	1.432
	Valence-band mean effective masses:	
m_{hh}^*	0.537 m_0	0.529 m_0
m_{lh}^*	0.157 m_0	0.157 m_0

To achieve better boundary conditions in nanostructures, Si-H parameters have been fitted on the SiH_4 experimental excitonic gap[20] and charge transfer calculated within LDA. H atoms are described by their 1s orbital. Si-H parameters are also reported in Table I.

B. Comparison between TB and $k\cdot p$ band structures

In this section, the TB band structure near the VBM and CBM is compared with a $\mathbf{k}\cdot\mathbf{p}$ description of bulk Si with consistent parameters. Since Si is an indirect band-gap semiconductor, we assume an uncoupled valence-band maximum and conduction band minima. The six uppermost valence bands (doubly degenerate heavy, light, and split-off bands) are described with a six- band $\mathbf{k}\cdot\mathbf{p}$ model (Dresselhaus-Kip-Kittel Hamiltonian[21]). It takes into account the large valence-band anisotropy of Si, and spin-orbit coupling. In this model, the periodic part $u_{n,\vec{k}}(\vec{r})$ of valence Bloch wave functions $\Psi_{n,\vec{k}}(\vec{r})=e^{i\vec{k}\vec{r}}u_{n,\vec{k}}(\vec{r})$ is expanded in the basis of the six uppermost valence Γ states $\{u_{m,\vec{k}=\vec{0}}(\vec{r})\}$. The input parameters for this model are the three Luttinger parameters γ_1,γ_2, and γ_3, and the spin-orbit splitting Δ. The six conduction-band minima along ΓX-like directions are assumed to be uncoupled from each other, and described by a single-band effective-mass approximation. The input parameters for the EMA are the longitudinal and transverse effective masses m_l^* and m_t^*, and band-gap energy E_g.

For consistent comparison, the TB Luttinger parameters and CB effective masses are used in the six-band $\mathbf{k}\cdot\mathbf{p}$ model and the EMA. The six-band $\mathbf{k}\cdot\mathbf{p}$ model and TB bulk valence bands are shown in Fig. 2, where we see that the six-band $\mathbf{k}\cdot\mathbf{p}$ model fails to reproduce long-range dispersion. The valence bands tend to acquire too much dispersion because they miss couplings with other states, which are not included in the six-band $\mathbf{k}\cdot\mathbf{p}$ model.[25] The quality of the bulk $\mathbf{k}\cdot\mathbf{p}$ valence bands strongly depends on the wave-vector direction. The mean difference between TB and $\mathbf{k}\cdot\mathbf{p}$ valence

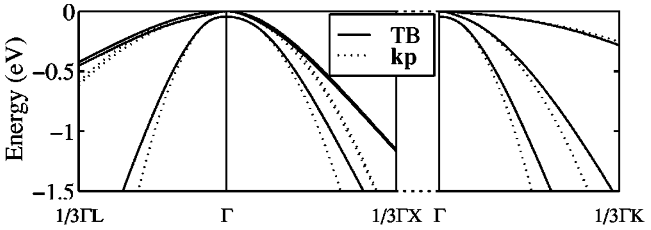

FIG. 2. Bulk Si valence-band structure within six-band **k**·**p** and TB models. TB Luttinger parameters are used for a consistent comparison. Spin-orbit coupling is taken into account.

bands is less than 10% in a ~250 meV range. In the same way, the lowest conduction band acquires too much dispersion around CBM's when compared to the TB model, especially in transverse mass directions. Also, the mean difference between TB and EMA conduction bands is less than 10% in a ~200-meV range.

III. APPLICATION OF THE TB MODEL TO Si NANOSTRUCTURES

The TB model is applied here to the calculation of the electronic structure of various free-standing Si nanostructures including Si(100) and (110) films, [100]- and [110]-oriented square-based Si wires and cylinders, and ''spherical'' (T_d symmetry) and cubic Si dots. Surface dangling bonds are saturated with hydrogen atoms. Spin-orbit coupling is taken into account, which was not the case in most previous calculations[7,9,10] (details of the TB calculation can be found in Appendix A). In all the cases, TB calculations are compared with six-band **k**·**p** (VB) or EMA (CB) results (details of the **k**·**p** and EMA calculations can be found in Appendix B).

Because the **k**·**p** model is not an atomistic description, there is no thorough way to provide a **k**·**p** potential consistent with TB boundary conditions. Thus an infinite barrier is assumed in **k**·**p** and EMA calculations, and its position is chosen in such a way that the volume of the system is equal to the total volume occupied by the Si atoms (see Appendix B for explicit expressions of the size in each case).

A. Confinement energies versus size

Results for Si(100) films are given in Fig. 3 versus film thickness L_f. The upper part of the figure shows the energy of the highest valence subband maximum at Γ (in-plane wave vector $\mathbf{k}_{\parallel}=0$) within the TB and six-band **k**·**p** models. The lower part shows the energy of the lowest and second conduction subband minima at Γ, as well as the energy of the other subband minima along [010] and [001] directions. Note that the lowest and second conduction subbands at Γ are degenerate in the EMA method, which is not the case in the TB method. They exhibit a large and pseudoperiodic splitting, due to intervalley coupling between opposite [100] bulk CB minima.[22]

Results for [100]-oriented wires with various shapes are shown in Fig. 4. Square wires with either $(010)\times(001)$ or $(011)\times(0\bar{1}1)$ faces and cylinders were considered. The energy of the highest valence and lowest conduction subbands at Γ (longitudinal wave vector $k_l=0$) is plotted versus the effective diameter d_1 of the wire, which is the diameter of the cylinder with the same transverse section as the wire.

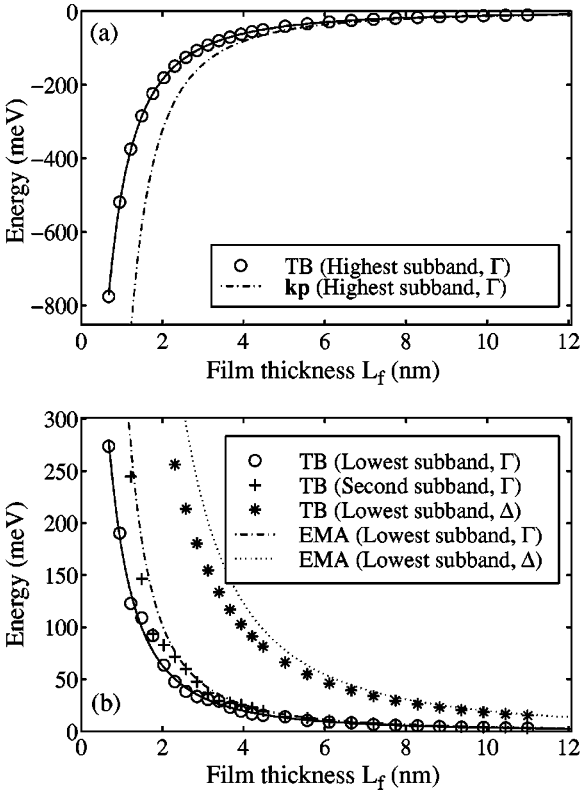

FIG. 3. (a) Energy of the highest valence subband maximum at Γ in Si (100) films. (b) Energy of the lowest and second conduction subband minima at Γ and energy of the other subband minima along [010] and [001] (Δ). Results are given vs film thickness L_f, within the TB and six-band **k**·**p** or EMA models. Energy is measured with respect to the bulk VBM (a) and CBM (b). The solid line is a fit to TB results. Spin-orbit coupling is taken into account.

Results for spherical and cubic $[(100)\times(010)\times(001)$ faces] Si dots are shown in Fig. 5. The energy of the highest occupied state (HOS) and lowest unoccupied state (LUS) are plotted versus the effective diameter d_0 of the dot, which is the diameter of the sphere with the same volume as the dot.

The energy of the TB highest valence subband maximum (or HOS) and lowest conduction subband minimum (or LUS) is fitted in the whole 1–12 nm range with the following expressions:

$$E_v(d)=\frac{K_v}{d^2+a_v d+b_v} \quad (\mathrm{HOS}), \tag{1}$$

$$E_c(d)=\frac{K_c}{d^2+a_c d+b_c}+E_g \quad (\mathrm{LUS}). \tag{2}$$

d is the characteristic dimension of the nanostructure; K,a, and b are adjustable constants; $E_g=1.143$ eV is the bulk band-gap energy. This expression is more accurate than the widely used fit K/d^{α} when a large range of dimensions is considered. It correctly behaves like $1/d^2$ in large structures, so that it can be considered as valid over the whole range of

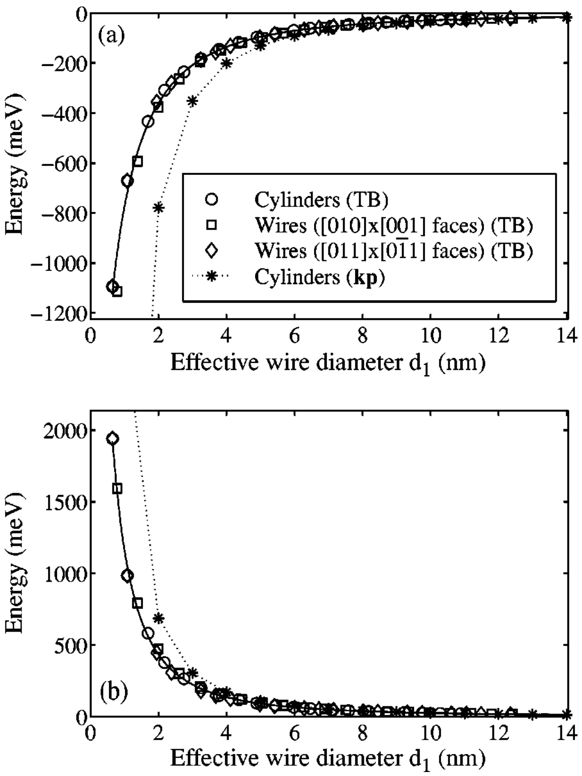

FIG. 4. Energy of the highest valence subband maximum (a) and lowest conduction subband minimum (b) at Γ for [100]-oriented Si wires with various shapes. Results are given vs the effective wire diameter d_1 (see text), within TB and **k**·**p** or EMA models. Also see Fig. 3 for more details.

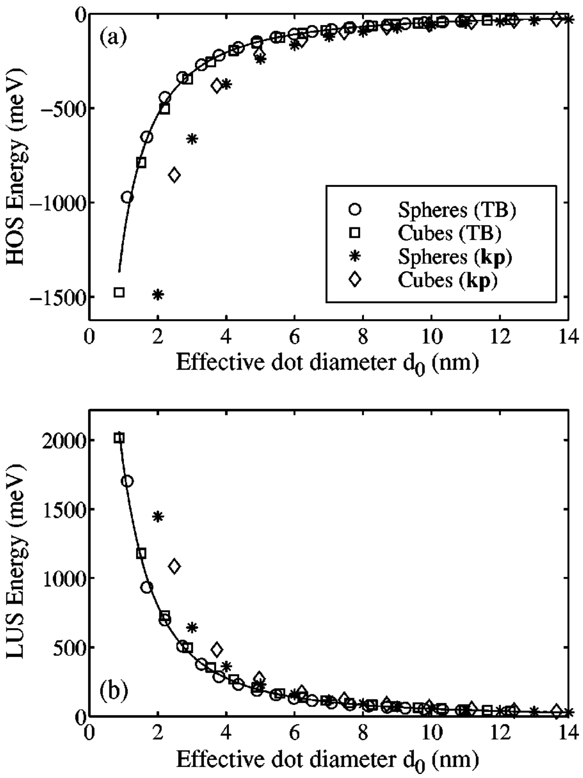

FIG. 5. Energy of the highest occupied state (a) and lowest unoccupied state (b) for spherical and cubic [(100)×(010)×(001) faces] Si dots. Results are given vs the effective dot diameter d_0 (see text), within TB and **k**·**p** or EMA models. Also see Fig. 3 for more details.

sizes. The fits are reported in Table III for all nanostructures considered in this work. They are also reported in Figs. 3, 4, and 5 (solid lines).

As shown in Fig. 4 for [100]-oriented wires and in Fig. 5 for dots, the TB band-gap energy mainly depends on the transverse section of the wires or volume of the dots as long as their shape is not too prolate.[7] Note, however, that this result cannot be extended to higher excited states.

Spin-orbit coupling has little influence on the highest valence subband maximum or HOS energy in Si(100) films, [100]-oriented wires, or spherical and cubic Si dots. Without spin-orbit coupling, the error on $E_v(d)$ is less than 10% in the whole 1–12-nm range, but increases monotonically with nanostructure size. Spin-orbit coupling has much more influence on the energy of the highest valence subband maximum of Si(110) films and [110]-oriented wires. It always significantly affects higher excited hole states, whatever the nanostructure. This will be discussed below for Si clusters.

Our TB parameters provide considerable improvement over previous sp^3 TB fits. This may be especially be shown in Si films, where conduction- and valence-band anisotropies must be accurately reproduced. Taking the example of Si(100) films, the third-nearest-neighbor sp^3 TB model of Ref. 15 gives nearly a 60% higher confinement energy for electrons ($m_t^*=0.567m_0$), but only half the confinement energy for holes ($\gamma_2=1.233$). This previous sp^3 TB fit cannot therefore be safely applied to Si films, though it gives quite equivalent band-gap energies in small spherical dots. Conversely, our fit provides accurate results for any Si nanostructure, whatever its dimensionality.

We now proceed to test our TB model against various other descriptions such as PP, LDA, or other TB models in the case of small spherical Si dots. Our results are first compared with those of Ref. 7 calculated with pseudopotentials in the range 1–4 nm. As shown in Fig. 6, they are in good agreement. Comparison with the *ab initio* LDA calculations of Ref. 5 corrected for the bulk band-gap error of 0.65 eV, shows that our results are extremely good down to the smallest clusters, except for some oscillations in the LDA that do not exist in TB. Finally, the agreement with other more complex TB models with overall band structures of similar quality, such as the $sp^3d^5s^*$ model of Ref. 16 or the nonorthogonal TB model of Ref. 14, is also excellent for small ($d_0<4$ nm) crystallites. As a consequence, the results previously published by our group[9,10] for small spherical Si dots remain valid.

B. Density of states in Si clusters

Figure 7 shows the ''valence'' (filled states) and ''conduction'' (empty states) densities of state, (DOS's) for a spherical Si dot with diameter $d_0=7.61$ nm, within TB and **k**·**p** or EMA models. Spin-orbit coupling is not taken into account in this figure. Energies are measured with respect to the bulk VBM for filled states, and to the bulk CBM for empty states. The states are labeled with the irreducible rep-

TABLE III. Fits to the energy of the highest valence subband maximum (or HOS) and lowest conduction subband minimum (or LUS) for various Si nanostructures, using expressions (1) and (2) (see text).

K_v (meV nm^2)	a_v (nm)	b_v (nm^2)	K_c (meV nm^2)	a_c (nm)	b_c (nm^2)
Si(100) films with thickness $d = L_f$:					
-1326.2	1.418	0.296	394.5	0.939	0.324
Si(110) films with thickness $d = L_f$:					
-1019.3	10.371	-3.713	1123.2	0.535	1.481
[100]-oriented Si cylinders with diameter $d = d_1$:					
-3448.4	2.194	1.386	2811.6	1.027	0.396
[100]-oriented Si wires with (010)×(001) faces and width $d = d_1\sqrt{\pi}/2$:					
-2817.9	1.988	0.708	2378.6	0.883	0.400
[100]-oriented Si wires with $(011)\times(0\bar{1}\bar{1})$ faces and width $d = d_1\sqrt{\pi}/2$:					
-2981.7	2.386	1.087	2162.0	1.129	0.138
[110]-oriented Si cylinders with diameter $d = d_1$:					
-2551.8	2.970	0.813	2860.8	1.330	2.650
[110]-oriented Si wires with $(0\bar{1}\bar{1})\times(001)$ faces and width $d = d_1\sqrt{\pi}/2$:					
-2217.5	3.177	0.130	2684.3	2.237	1.408
Spherical Si dots with diameter $d = d_0$:					
-6234.0	3.391	1.412	5844.5	1.274	0.905
Cubic Si dots with (100)×(010)×(001) faces and side $d = d_0(\pi/6)^{1/3}$:					
-3967.0	2.418	0.522	4401.0	1.138	0.889

resentations of the T_d group. Figure 8 shows the valence TB DOS in the same dot, with spin-orbit coupling taken into account for comparison.

Without spin-orbit coupling, the T_2 HOS is *s* like (no nodes), with protrusions along {111}-like directions. The next occupied T_1 states are *p* like, with a nodal plane. Both T_2 and T_1 states can accommodate six electrons (threefold space and twofold spin degeneracy). These sixfold-degenerate states are split by spin-orbit coupling into one fourfold- and one twofold-degenerate states. Further coupling between split states may strongly affect the DOS, depending on the size of the cluster. When Δ is much higher than the splitting between T_2 and T_1 states ($d_0 > 5$ nm in spherical Si dots), the HOS and next occupied states remain *s* and *p* like, but can now accommodate only four electrons (see Fig. 8). The HOS confinement energy remains nearly unchanged, but the higher excited spectrum is significantly modified. When Δ is much lower than the splitting between T_2 and T_1 states ($d_0 < 5$ nm), the HOS is fourfold degenerate *s* like, followed by the other twofold-degenerate *s*-like state, then by *p*-like states. Although spin-orbit coupling has a negligible influence on the HOS energy, it may thus affect tunneling spectroscopy and Coulomb blockade experiments in Si clusters.

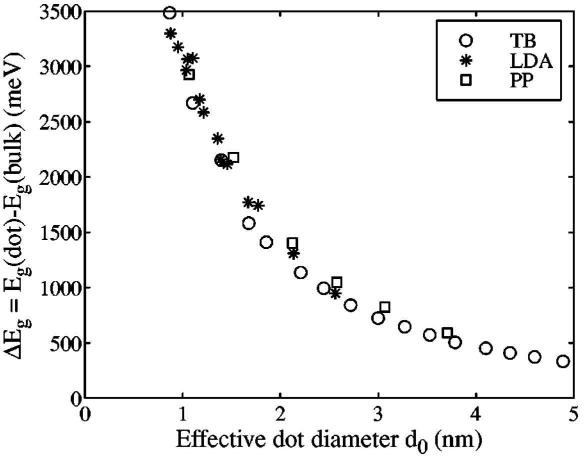

FIG. 6. Comparison of confinement energy $\Delta E_g = E_g(\text{dot}) - E_g(\text{bulk})$ in spherical Si dots, between our TB model and pseudopotential (PP) or LDA.

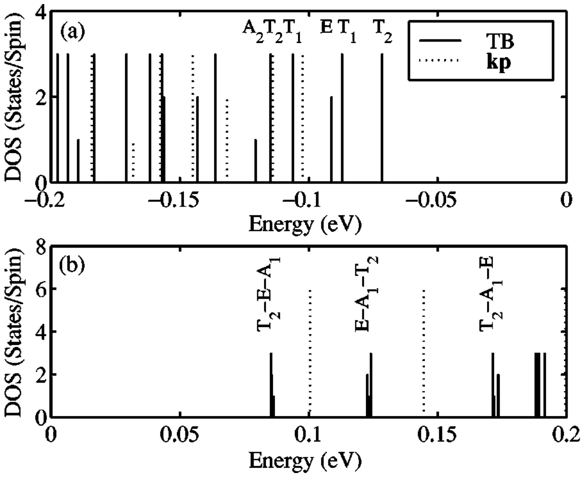

FIG. 7. ''Valence band'' [filled states (a)] and ''conduction band'' [empty states (b)] DOS is for a spherical Si dot with diameter $d_0 = 7.61$ nm, within TB and $\mathbf{k}\cdot\mathbf{p}$ or EMA models. Energy is measured with respect to the bulk VBM (upper part) and CBM (lower part). Spin-orbit coupling is not taken into account. The states are labeled with irreducible representations of the T_d group.

The splitting between T_2 and T_1 states increases when decreasing γ_2 or γ_3. Again, a correct description of valence-band anisotropy is needed in the computation of the hole states. In particular, we do not observe any change in the HOS symmetry (T_2) in the whole 1–12-nm range.[11]

The EMA LUS is sixfold degenerate, since the bulk CB minima are assumed uncoupled. Their envelope functions are elongated ellipsoids oriented along the longitudinal

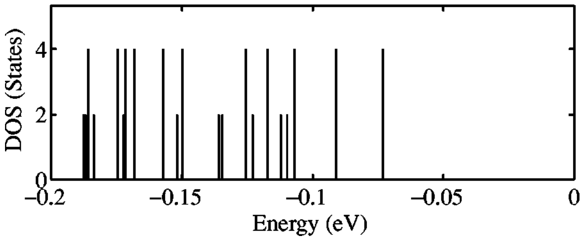

FIG. 8. ''Valence-band'' DOS for the same spherical Si dot as in Fig. 7 ($d_0 = 7.61$ nm). Spin-orbit coupling is taken into account.

(heavy) mass direction of each of the six bulk CB minima. These sixfold-degenerate EMA states are split into groups of one A_1 state, one twofold-degenerate E state, and one threefold-degenerate T_2 state in the TB model by intervalley couplings. The ordering of these states in each group is not the same from dot to dot. The splittings of the lowest-lying group in spherical Si dots range from 0.1 meV for diameter $d_0 = 10.32$ nm up to 68 meV for diameter $d_0 = 1.85$ nm.

IV. COMPARISON BETWEEN TB AND k·p MODELS IN SI NANOSTRUCTURES

Detailed comparisons between $\mathbf{k}\cdot\mathbf{p}$ and pseudopotential models have been performed in III-V (InP) and II-VI (CdSe) free-standing quantum dots,[23–25] GaAs/AlAs quantum wells,[26] and strained InAs dots embedded in GaAs.[27] In particular, these studies have shown the failure of the six-band $\mathbf{k}\cdot\mathbf{p}$ model in free-standing InP dots, leading to the interchange of *s*- and *p*-like valence states. Although $\mathbf{k}\cdot\mathbf{p}$ models are known to overestimate confinement in small Si nanostructures,[7] no comparison between $\mathbf{k}\cdot\mathbf{p}$ and TB or pseudopotential methods has been made in large, free-standing Si clusters.

There is striking evidence for the overconfinement of $\mathbf{k}\cdot\mathbf{p}$ and EMA models with respect to the model TB in Figs. 3, 4, and 5, even in large nanostructures, as evidenced by Fig. 7. $\mathbf{k}\cdot\mathbf{p}$ predictions obviously become worse from films to wires and dots. In spherical Si dots, the error on the $\mathbf{k}\cdot\mathbf{p}$ confinement energy $\Delta E_g = E_g(\text{dot}) - E_g(\text{bulk})$ is larger than 25% for $d_0 < 8.5$ nm, and 50% for $d_0 < 4.5$ nm. It is still 15% for $d_0 \simeq 12$ nm. The use of a multiband semiempirical method such as TB is therefore to be recommended in the 5–12-nm range for Si clusters. Indeed, the sp^3 TB model is not more difficult to solve than the six-band $\mathbf{k}\cdot\mathbf{p}$ model in this range, when valence-band anisotropy and spin-orbit coupling are taken into account.

Despite this clear quantitative disagreement between TB and $\mathbf{k}\cdot\mathbf{p}$ models, there is however qualitative correspondence in the density of states and wave-function ordering in Si clusters, as evidenced by Fig. 7. The few highest occupied and lowest empty $\mathbf{k}\cdot\mathbf{p}$ states can be associated with a TB state with consistent wave-function symmetry. However, the splitting between successive states is not consistent with TB calculations, the energy of higher excited states being further and further overestimated. The same conclusions may be drawn in large films and wires: the $\mathbf{k}\cdot\mathbf{p}$ models's highest valence and lowest conduction subband ordering is consistent with TB calculations, but they suffer from increasing overconfinement.

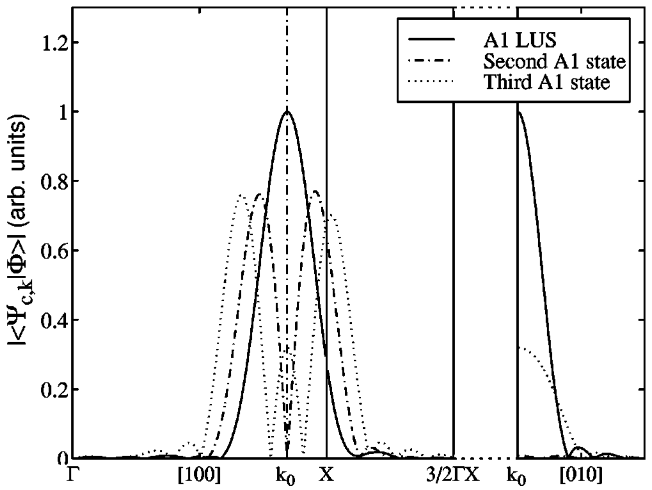

FIG. 9. Bloch decomposition of the three lowest empty A_1 states of a spherical dot with diameter $d_0 = 4.89$ nm. The decomposition is performed ΓX on the lowest two conduction bands. For clarity, $|\overline{\Phi}_{n,\vec{k}}|$ is shown in an extended zone scheme, the first conduction band being on the left on the X point, and the second one on the right (left part). $|\overline{\Phi}_{n,\vec{k}}|$ is also calculated on the first conduction band only in the transverse mass direction (right part).

We now discuss the reasons why the $\mathbf{k}\cdot\mathbf{p}$ (EMA) method overestimates confinement energies in Si nanostructures. We will focus on spherical Si dots. For the sake of simplicity, spin-orbit coupling is not taken into account in the following discussion. However, we have checked that our conclusions did not change if spin-orbit coupling was included. We will relate $\mathbf{k}\cdot\mathbf{p}$ errors in nanostructures to $\mathbf{k}\cdot\mathbf{p}$ errors in bulk description and boundary conditions. We thus introduce the ''Bloch decomposition'' of any nanostructure state Φ, which is the projection $\overline{\Phi}_{n,\vec{k}}$ of Φ on bulk Bloch wavefunctions $\Psi_{n,\vec{k}}$:[11,25]

$$\overline{\Phi}_{n,\vec{k}} = \langle \Psi_{n,\vec{k}} | \Phi \rangle. \tag{3}$$

$\Psi_{n,\vec{k}}(\vec{r}) = e^{i\vec{k}\vec{r}} u_{n,\vec{k}}(\vec{r})$ is the Bloch wave function with wave vector $\mathbf{k}$ and band index n. This Bloch decomposition is performed on the restriction of the TB wave functions Φ to the Si part of the cluster, thus excluding H atoms.

There are three reasons why $\mathbf{k}\cdot\mathbf{p}$ and EMA methods overestimate confinement energies with respect to TB. The main reason is the poor bulk description within the $\mathbf{k}\cdot\mathbf{p}$ model.[25] This was discussed in Sec. II B: valence and conduction bands tend to acquire too much dispersion far from the VBM and CBM. $\mathbf{k}\cdot\mathbf{p}$ and EMA methods will thus overestimate confinement energies in nanostructures. Overconfinement becomes worse in small nanostructures that couple Bloch states farther and farther away from the VBM and CBM, and for higher excited states. As an example, $|\overline{\Phi}_{n,\vec{k}}|$ is shown in Fig. 9 for the three lowest empty A_1 states of a spherical Si dot with diameter $d_0 = 4.89$ nm. The decomposition is performed along ΓX on the lowest two conduction bands (left part of Fig. 9). For clarity, $|\overline{\Phi}_{n,\vec{k}}|$ is shown in an extended zone scheme, the first conduction band being on the left on the X point, and the second one on the right. $|\overline{\Phi}_{n,\vec{k}}|$ is also calculated on the first conduction band only in the transverse mass direction (right part of Fig. 9). The full width at half maximum of the main peak of the A_1 LUS is proportional to $1/d_0$, and extends in the whole Brillouin zone in the smallest nanostructures. Higher excited states, that have nodal planes in the wave function, thus exhibit multiple peaks that extend farther in reciprocal space, beyond the range of validity of bulk $\mathbf{k}\cdot\mathbf{p}$ and EMA descriptions.

The next reason is the coupling between bulk bands in nanostructures.[25] Indeed, the six-band $\mathbf{k}\cdot\mathbf{p}$ model assumes that filled states can be decomposed on the six highest bulk valence bands. TB calculations, however, show that the highest occupied states have nonzero projections on bulk conduc-

TABLE IV. Total projection on bulk valence bands P_v, conduction bands P_c, and hydrogens P_H for the HOS and LUS of spherical Si dots with diameters $d_0 = 2.44$ nm and 7.61 nm. The corresponding contributions E_v,E_c, and E_H to the HOS and LUS energy E are also given.

		$d_0 = 2.44$ nm	$d_0 = 7.61$ nm
HOS	E (eV)	-0.414	-0.086
	P_v (%)	88.44	98.89
	E_v (eV)	-0.580	-0.108
	P_c (%)	4.47	0.44
	E_c (eV)	0.237	0.024
	P_H (%)	7.09	0.66
	E_H (eV)	-0.071	-0.002
LUS	E (eV)	1.738	1.228
	P_c (%)	95.94	99.71
	E_c (eV)	1.740	1.230
	P_v (%)	1.69	0.11
	E_v (eV)	-0.080	-0.006
	P_H (%)	2.37	0.18
	E_H (eV)	0.078	0.004

tion bands (see below). In the same way, the lowest empty states have nonzero projections on bulk valence bands and higher conduction bands. The details of interband coupling in nanostructures depend on boundary conditions. The trends are, however, universal: interband coupling increases with decreasing nanostructure size.

The last reason is the lack of correct boundary conditions in free-standing $\mathbf{k}\cdot\mathbf{p}$ nanostructures. Indeed, the $\mathbf{k}\cdot\mathbf{p}$ model cannot handle atomisticlike boundary conditions (e.g., a Si-H bond). However, the HOS and LUS wave functions may be partly delocalized over H atoms in small dots (see below). Therefore, hydrogen atoms will contribute to the confinement energy.

Results regarding interband coupling and boundary conditions are reported in Table IV. The total squared projection on bulk valence bands P_v, bulk conduction bands P_c, and hydrogen atoms $P_H (P_v + P_c + P_H = 1)$ is given for the HOS and LUS of spherical Si dots with diameters $d_0 = 2.44$ and 7.61 nm. The corresponding contributions E_v,E_c, and E_H to the HOS and LUS energy E are also given $(E_v + E_c + E_H = E)$. The coupling of the HOS with bulk conduction bands and hydrogen atoms is very important in the smallest dot $(d_0 = 2.44$ nm$)$. Although P_c is only 4.47%, the energy distributed over bulk conduction bands is as large as $E_c = +237$ meV $(E = -414$ meV$)$. The coupling of the HOS with conduction bands remains significant for $d_0 = 7.61$ nm, where still $E_c = +24$ meV $(E = -86$ meV$)$. The coupling of the LUS with bulk valence bands is less significant, but cannot be neglected either.

The only way to improve the bulk $\mathbf{k}\cdot\mathbf{p}$ description and allow interband coupling is to increase the number of bands in the $\mathbf{k}\cdot\mathbf{p}$ model.[28] However, such a procedure is not really efficient, especially in indirect band-gap materials. Although the deficiencies of the $\mathbf{k}\cdot\mathbf{p}$ models are much more sensitive in free-standing nanostructures than in heterostructures, they will be important each time silicon nanostructures are highly confined, as for Si nanocristallites embedded in SiO_2. In the latter case, only TB or PP models can properly handle Si-O bonding, and should clearly be preferred to $\mathbf{k}\cdot\mathbf{p}$ models.

V. CONCLUSION

We have introduced an orthogonal third-nearest-neighbor TB model for silicon. Special emphasis was given in the description of the valence-band maximum and conduction-band minima, including conduction- and valence-band effective masses in the TB fit. This model was applied to the calculation of the edge states of various Si nanostructures (films, wires, and dots), including spin-orbit coupling, over the whole 1–12-nm range. Analytical laws for the confinement energies were given for the purpose of comparison with experiment or with other models. The comparison with the LDA, pseudopotential and $sp^3d^5s^*$ models was performed to test the accuracy of this TB fit. We showed improvement in the electronic structure of Si nanostructures with respect to previous sp^3 TB models. The deficiencies of $\mathbf{k}\cdot\mathbf{p}$ and EMA models in free-standing Si nanostructures were discussed and quantified with the TB model. We showed that the $\mathbf{k}\cdot\mathbf{p}$ makes substantial errors even in large zero-dimensional nanostructures, due to a poor bulk description and lack of interband coupling. The use of this efficient and accurate TB model in the calculation of transport properties of Si nanostructures should allow a comprehensive study of the physics of these devices.

ACKNOWLEDGMENTS

We thank Lucia Reining for providing an accurate GW band structure of silicon. The Institut d’Electronique et de Microélectronique du Nord is UMR 8520 of CNRS.

APPENDIX A: TB CALCULATION OF THE ELECTRONIC STRUCTURE OF Si NANOSTRUCTURES

The tight-binding model introduced in Sec. II can be applied to any Si nanostructure, once given the atomic positions. To avoid surface states in the gap, all dangling bonds are saturated with hydrogen atoms. Spin-orbit coupling is directly taken into account in Si films and wires using a spin-augmented sp^3 basis. However, spin-orbit coupling doubles the dimension of the Hamiltonian, and makes it complex Hermitian, rather than real symmetric in dots. Being small in Si, it is calculated *a posteriori* in dots in the basis of the few highest occupied states of the spin-orbit free-dot Hamiltonian. About ten states are usually enough to ensure convergence of the hole ground-state energy. Spin-orbit coupling has negligible influence on the lowest conduction states.

In quantum dots, where no translationnal symmetries can help reduce the size of the problem, a basis is first constructed for each of the five irreducible representations of the T_d group. According to Wigner’s theorem,[33] this leads to a block-diagonal Hamiltonian with one A_1 block, one A_2 block, two equivalent E blocks, three equivalent T_1 blocks, and three equivalent T_2 blocks. One block is then processed separately for each representation, the eigenstates of the other degenerate equivalent blocks being computed from the latter. Hundreds of valence or conduction states can then be computed in a reasonable time.

According to the dimension d of the Hamiltonian or its blocks, either all eigenpairs are computed using a standard QR algorithm,[30] or only a few near the gap using a block Lanczos algorithm[30,31] on $(H-\sigma I)^{-1}$ or a conjugate gradients algorithm[32] on $(H-\sigma I)^2$ (folded spectrum method[6]). The folding energy σ is set in the gap just above the bulk valence-band maximum or below the bulk conduction-band minimum to directly catch the highest valence or lowest conduction states. Jacobi (diagonal) or incomplete Cholesky factorizations[30] $(LL^{\dagger})$ of $(H-\sigma I)^2$ were used as preconditioners for the conjugate gradients. In the latter case, the incomplete Cholesky factorization was performed on the part of $(H-\sigma I)^2$ having the sparsity pattern of H. Although crude, this preconditionner can save up to 75% of the iterations needed to reach convergence depending on the problem.

APPENDIX B: k·p AND EMA CALCULATIONS OF THE ELECTRONIC STRUCTURE OF Si NANOSTRUCTURES

Tight-binding calculations are compared with effective-mass or $\mathbf{k}\cdot\mathbf{p}$ approximations.[29] A full six-band $\mathbf{k}\cdot\mathbf{p}$ model, including spin-orbit coupling, is used in the valence band (Dresselhaus-Kip-Kittel Hamiltonian.[21]) The six conduction-band minima are assumed to be uncoupled, and treated in the single-band anisotropic effective-mass approximation. Envelope functions are expanded on a finite-elements,[34] tensor product mesh with linear interpolation functions. The generalized eigenvalue problem is then solved with a conjugate gradients algorithm[35] preconditionned with an incomplete Cholesky $(LDL^{\dagger})$ factorization with zero fill.

An infinite barrier is assumed in $\mathbf{k}\cdot\mathbf{p}$ (EMA) calculations. Its position is chosen in such a way that the volume of the system is equal to the total volume occupied by the N_{Si} Si atoms. Furthermore, to allow comparison between clusters with different shapes, an effective diameter d_0 is defined for any cluster as the diameter of the sphere with the same volume as the cluster. Thus we obtain

$$d_0 = a\left(\frac{3}{4\pi}N_{\mathrm{Si}}\right)^{1/3} = 0.33691 N_{\mathrm{Si}}^{1/3}(\mathrm{nm}). \tag{B1}$$

In the same way, an effective diameter d_1 is defined for any wire as the diameter of the cylinder with the same transverse section as the wire. In [100]- and [110]- oriented Si wires we get $d_1 = 0.21667 N_{\mathrm{Si}}^{1/2}$ (nm) and $d_1 = 0.25766 N_{\mathrm{Si}}^{1/2}$ (nm) where N_{Si} is the number of Si atoms in the $a[100]$ and $a\sqrt{2}/2[110]$ supercells. Finally, the thickness of Si(100) and (110) films is $L_f = N_{(100)}a/4 = 0.13578 N_{(100)}$ and $L_f = N_{(110)}a/(2\sqrt{2}) = 0.19201 N_{(110)}$, respectively, where $N_{(100)}$ and $N_{(110)}$ are the number of Si planes in the film.

---

*Corresponding author: Y. M. Niquet.
E-mail: ymn@isen.iemn.univ-lille1.fr
FAX: (33) 3 20 19 78 84.

[1] A.C. Seabaugh, A.H. Taddiken, E.A. Beam, J.N. Randall, and Y.C. Kao, Electron. Lett. **29**, 1802 (1993).

[2] E. Leobandung, L. Guo, Y. Wang, and S.Y. Chou, Appl. Phys. Lett. **67**, 938 (1995).

[3] A. Fujiwara, Y. Takahashi, K. Murase, and M. Tabbe, Appl. Phys. Lett. **67**, 2957 (1995).

[4] *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures*, edited by H. Grabert and M. Devoret, Proceedings of a NATO Advanced Study Institute on Single Charge Tunneling held in Les Houches, France (1991) (Plenum Press, New York, 1992).

[5] B. Delley and F. Steigmeier, Appl. Phys. Lett. **67**, 2370 (1995).

[6] Lin-Wang Wang and A. Zunger J. Chem. Phys. **100**, 2394 (1994).

[7] A. Zunger and Lin-Wang Wang, Appl. Surf. Sci. **102**, 350 (1996).

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

[9] J.P. Proot, C. Delerue, and G. Allan, Appl. Phys. Lett. **61**, 1948 (1992).

[10] C. Delerue, G. Allan, and M. Lannoo, Phys. Rev. B **48**, 11 024 (1993).

[11] Shang Yuan Ren, Phys. Rev. B **55**, 4665 (1997).

[12] T. Takagahara and K. Takeda, Phys. Rev. B **46**, 15 578 (1992).

[13] P. Vogl, H.P. Hjalmarson, and J. Dow, J. Phys. Chem. Solids **44**, 365 (1983).

[14] P.B. Allen, J.Q. Broughton, and A.K. MacMahan, Phys. Rev. B **34**, 859 (1986).

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

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

[17] O. Madelung, *Semiconducters Physics of Group IV Elements and III-V Compounds*, edited by Landölt-Bornstein, New Series, Group III, Vol. 17, Pt. a (Springer-Verlag, Berlin, 1992).

[18] M.S. Hybertsen and S.G. Louie, Phys. Rev. Lett. **55**, 1418 (1985).

[19] L. Reining (private communication).

[20] M. Rholfing and S.G. Louie, Phys. Rev. Lett. **80**, 3320 (1998).

[21] G. Dresselhaus, A.F. Kip, and C. Kittel, Phys. Rev. **98**, 368 (1955).

[22] T. Ando, A.B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).

[23] Huaxiang Fu, Lin-Wang Wang, and A. Zunger, Appl. Phys. Lett. **71**, 3433 (1997).

[24] Huaxiang Fu, Lin-Wang Wang, and A. Zunger, Appl. Phys. Lett. **73**, 1157 (1998).

[25] Huaxiang Fu, Lin-Wang Wang, and A. Zunger, Phys. Rev. B **57**, 9971 (1998).

[26] D.M. Wood and A. Zunger, Phys. Rev. B **53**, 7949 (1996).

[27] Lin-Wang Wang, J. Kim, and A. Zunger, Phys. Rev. B **59**, 5678 (1999).

[28] Lin-Wang Wang and A. Zunger, Phys. Rev. B **54**, 11 417 (1996).

[29] D. Gershoni, C.H. Henry, and G.A. Baraff, IEEE J. Quantum Electron. **29**, 2433 (1993).

[30] G. H. Golub, *Matrix Computations*, 3rd ed. (Johns Hopkins University Press, Baltimore, 1996).

[31] C. Lanczos, J. Res. Natl. Bur. Stand. **45**, 255 (1950).

[32] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias, and J.D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).

[33] Wu-Ki Tung, *Group Theory in Physics* (World Scientific, Philadelphia, 1985).

[34] O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method* (McGraw-Hill, New York, 1989).

[35] Yang Haigang, Ph.D. thesis, Philosophische Fakultät Zürich, 1993.