

Self-interaction corrections in semiconductors

Martin M. Rieger* and P. Vogl

*Walter Schottky Institut und Physik Department der Technischen Universität München,
Am Coulombwall, 85748 Garching, Germany*

(Received 30 March 1995)

We propose a scheme to generate pseudopotentials from strictly self-consistent self-interaction-corrected local-density-approximation (SIC-LDA) calculations, that are firmly rooted within density functional theory, and use them in a standard LDA calculation of the structural properties and the eigenvalue spectrum in crystalline Si, Ge, Sn, and GaAs. Thus we avoid the serious shortcomings of LDA in small atoms and nevertheless use consistently the same method in both atoms and solids, since LDA and SIC-LDA coincide for the completely delocalized valence electrons of these materials. We show that significant effects arise from an improved description of the core. In particular, Ge, which appears to be a metal with band overlap at the Γ point of the Brillouin zone in LDA, can be correctly described as an indirect-gap semiconductor. Still, it is argued that fully self-consistent self-interaction corrections can only eliminate a small part of the gap problem in *sp*-bonded semiconductors.

I. INTRODUCTION

Density-functional methods have been of paramount prominence in the field of *ab initio* calculations of the electronic structure of solids for the past 20 years. Rooted in the theorem of Hohenberg and Kohn,¹ it was the local-density approximation (LDA), together with the Kohn-Sham scheme,² that opened the way for practical calculations. The LDA is, by construction, exact in the homogeneous electron gas. In nonuniform systems, the problem of possible unphysical self-interactions comes up, as each electron contributes to the total density and thus to an effective potential that acts again on the same electron. This problem becomes particularly serious in small atoms, where two important properties of the exact density functional are violated by LDA, namely the long-range behavior of the exchange-correlation potential, which should decay³ like $1/r$, and the equivalence of the highest occupied Kohn-Sham eigenvalue and the first ionization potential.⁴

These self-interactions can be explicitly corrected for and, although this does not lead to the elusive "exact" exchange-correlation functional, at least the proper long-range behaviour of the effective potential can be restored by this physically motivated approximation and the gap between first ionization potential and highest eigenvalue can be reduced for the largest part. Self-interaction corrections to the local-density approximation or its spin-polarized variant, the local spin-density approximation⁵ (LSD), have been successfully applied to atoms and small molecules⁶⁻¹⁰ as well as to extended systems.¹¹⁻²⁰ The formalism has its roots in the work of Perdew and Zunger⁶ and has since been developed further by other scientists in the field (see the above citations). The occurrence in this formalism of an apparently state-dependent potential has led to the impression that it is outside the domain of density-functional the-

ory. We have proposed²¹ recently a scheme for a relativistic self-interaction-corrected LDA (SIC-LDA) which we could show to be firmly rooted in density-functional theory and have applied it to free atoms, where it yields results considerably improved over the normal LDA.

In this paper we investigate how self-interaction corrections that follow strictly the demands of self-consistency — albeit within the constraints set by the *ab initio* pseudopotential method and the rigid core approximation — affect the electronic and structural properties of semiconductors in first-principles calculations. The valence electrons of ideal semiconductor crystals are well described by Bloch functions. By their very nature, self-interaction corrections arise only from states that are localized within a finite volume of space. Therefore any possible effects that can be brought about by SIC in these types of materials rely on an improved description of the localized core electrons.

Most *ab initio* calculations for semiconductors use the pseudopotential formalism and here again the overwhelming majority of pseudopotentials is generated from LDA calculations in atoms (e.g., Refs. 22–25), which are rather inadequate, as explained above. The reason why, nevertheless, the LDA is used in the generation of pseudopotentials is to be seen in the fact that pseudopotential theory requires the same method to be employed both in the construction of a pseudopotential and in its subsequent application. Our approach is to apply consistently SIC-LDA in the atom and the crystal. In the former, SIC leads to striking improvements while in the latter, SIC-LDA and LDA coincide. The better description of the atomic core is transmitted to the solid by the pseudopotential generated from atomic SIC-LDA calculations.

Pseudopotentials with some approximate self-interaction corrections have been proposed earlier by Zunger²⁶ and Shirley *et al.*²⁷ and tested in atoms and small molecules. The present work constitutes, to the best of

our knowledge, the first attempt to generate pseudopotentials from fully self-consistent relativistic SIC-LDA calculations and apply them in infinite solids.

One set of quantities we will be looking at are the band gaps. The latter are traditionally very poorly described in LDA and constitute the most visible failure of the LDA in semiconductors. Density-functional theory, albeit a ground-state theory, is capable in principle to provide the fundamental gap in a semiconductor crystal as the difference of the Kohn-Sham eigenvalues of the highest occupied states of an N - and $(N + 1)$ -electron system:²⁸

$$\epsilon_g = \epsilon_{N+1}(N + 1) - \epsilon_N(N). \quad (1.1)$$

Perdew and Levy²⁹ and Sham and Schlüter³⁰ could show that this gap is connected to the Kohn-Sham eigenvalues of the N -electron system by the relation

$$\epsilon_g = \epsilon_c - \epsilon_v + \Delta_{xc}, \quad (1.2)$$

where ϵ_c is the Kohn-Sham eigenvalue of the bottom of the conduction band and ϵ_v that of the top of the valence band. Δ_{xc} denotes a discontinuity in the functional derivative of the exchange-correlation energy by the occupation numbers.

The LDA possesses no such discontinuity,

$$\Delta_{xc}^{\text{LDA}} = 0, \quad (1.3)$$

so that it is consistent within the LDA model to interpret the eigenvalue gap as the fundamental gap. However, the model itself suffers from deficiencies, as witnessed by the poor gaps in semiconductors. It is not quite clear, though, how much of the error is to be attributed to the lacking discontinuity of the LDA exchange-correlation energy.³¹⁻³⁴ The SIC-LDA formalism for semiconductors we develop here will eventually not lead to a discontinuous exchange-correlation energy derivative and thus we cannot, with this method, fully eliminate the gap problem. However, we do show that significant gains in the description of the valence electrons can be achieved

through self-interaction corrections to the core alone.

The paper is organized as follows. In the second section we describe the principles of SIC-LDA, in the third the construction of the pseudopotentials is laid out, and in the fourth numerical results are presented and discussed. A summary concludes the paper. Note that all equations are written in atomic units.

II. SELF-INTERACTION CORRECTIONS

In this section we repeat the key elements of the SIC-LDA formalism.²¹ A nonrelativistic picture is chosen for greater clarity.

The SIC-LDA energy functional of an N -electron system to be minimized with respect to the density is in full generality

$$E^{\text{SIC-LDA}}[\rho] = \Theta^{\text{LDA}}[\rho] + U_H[\rho] + E_{xc}^{\text{LDA}}[\rho] + \int \rho(\mathbf{r})v_{\text{ext}}(\mathbf{r})d^3r. \quad (2.1)$$

Here, ρ denotes the density, $U_H[\rho]$ the Hartree energy, E_{xc}^{LDA} the local-density approximation to the exchange-correlation energy, v_{ext} any external potential, and

$$\Theta^{\text{LDA}}[\rho] = \min_{\{\phi_i\} \rightarrow \rho} \sum_{i=1}^N \left\{ \langle \phi_i | T | \phi_i \rangle - U_H[|\phi_i|^2] - E_{xc}^{\text{LDA}}[|\phi_i|^2] \right\}. \quad (2.2)$$

The minimization in Eq. (2.2) is over all sets of orthonormal one-electron wave functions ϕ_i that fulfill

$$\sum_{i=1}^N |\phi_i| = \rho, \quad (2.3)$$

and T is the kinetic energy operator.

Instead of minimizing $E^{\text{SIC-LDA}}[\rho]$ with respect to the density, one can minimize the functional

$$E_{\Psi}^{\text{SIC-LDA}}[\{\Psi_i\}] = \sum_{i=1}^N \langle \Psi_i | T | \Psi_i \rangle + U_H[\rho] + E_{xc}^{\text{LDA}}[\rho] + \int \rho(\mathbf{r})v_{\text{ext}}(\mathbf{r})d^3r + \min_{\mathbf{M}} \sum_{i=1}^N \left(-U_H \left[\left| \sum_j M_{ji}^* \Psi_j \right|^2 \right] - E_{xc}^{\text{LDA}} \left[\left| \sum_j M_{ji}^* \Psi_j \right|^2 \right] \right) \quad (2.4)$$

with respect to the orthogonal one-electron wave functions Ψ_i . For each set of functions $\{\Psi_i\}$ one has to find among all unitary matrices \mathbf{M} the matrix \mathcal{M} that minimizes the second sum in Eq. (2.4). The minimum of the total functional is reached, if the equations

$$\left(T + v_H[\rho] + v_{xc}[\rho] + v_{\text{ext}} + \bar{v}^{\text{SIC}} + \sum_k^{\text{occ}} (v^{\text{SIC}}[\rho_k] - \bar{v}^{\text{SIC}}) |\phi_k\rangle \langle \phi_k| \right) |\Psi_i\rangle = \epsilon_i |\Psi_i\rangle \quad (2.5)$$

hold simultaneously for $i = 1, \dots, N$. The following symbols have been used:

$$\phi_k = \sum_j \mathcal{M}_{jk}^* \Psi_j, \quad (2.6)$$

$$\rho_k = |\phi_k|^2, \quad (2.7)$$

$$v^{\text{SIC}}[\rho_k] = -v_H[\rho_k] - v_{xc}[\rho_k], \quad (2.8)$$

$$\bar{v}^{\text{SIC}} = \frac{1}{N} \sum_k v^{\text{SIC}}[\rho_k], \quad (2.9)$$

where “occ” indicates that the summation is over occupied orbitals only. v_H and v_{xc} are the Hartree and exchange-correlation potentials. The orbitals Ψ_i are called “canonical orbitals” and represent the relevant wave functions in a one-electron picture,²¹ and the ϕ_i are called “localized orbitals.”

The set $\{\Psi_i^0\}$ which minimizes (2.4) yields the density $\rho_0 = \sum_{i=1}^N |\Psi_i^0|^2$ which minimizes the functional (2.1) and the equality

$$E^{\text{SIC-LDA}}[\rho_0] = E_{\Psi}^{\text{SIC-LDA}}[\{\Psi_i^0\}] \quad (2.10)$$

holds. Thus the equivalence of both minimizations is guaranteed.

The second sum in Eq. (2.4),

$$U_{\text{SIC}} = \min_{\mathcal{M}} \sum_{i=1}^N \left(-U_H \left[\left| \sum_j M_{ji}^* \Psi_j \right|^2 \right] - E_{xc}^{\text{LDA}} \left[\left| \sum_j M_{ji}^* \Psi_j \right|^2 \right] \right), \quad (2.11)$$

represents the self-interaction correction to the total energy and only this term distinguishes $E_{\Psi}^{\text{SIC-LDA}}$ from its LDA counterpart.

The critical problem in truly self-consistent SIC-LDA is to find the transformation \mathcal{M} between canonical and localized orbitals. In atoms, this problem has been solved.^{10,21} In solids it is rather more complicated. We therefore make a few assumptions for the materials we deal with in this paper. The first is that the transformation will not mix core and valence states. This is motivated by the relatively small overlap between core and valence states. Consequently U_{SIC} splits up into

$$U_{\text{SIC}}[\{\phi\}] = U_{\text{SIC}}[\{\phi_c\}] + U_{\text{SIC}}[\{\phi_v\}], \quad (2.12)$$

where the indices c and v denote core and valence functions.

The core states can be treated as atomiclike and the proper transformation is easily found (see the next section). The canonical valence wave functions of periodic semiconductor crystals, on the other hand, can be well represented by Bloch functions, which are extended over the full volume of the crystal. If we assume that for the valence states the canonical functions Ψ_v and the localized functions ϕ_v are identical, the valence part $U_{\text{SIC}}^v = U_{\text{SIC}}[\{\phi_v\}]$ of the self-interaction correction will be zero. If N_v valence electrons in a system of volume V are extended over the entire volume, the Hartree part of U_{SIC}^v scales like N_v/V^2 and the exchange-correlation part like $N_v/V^{4/3}$, so that at a fixed average density N_v/V it vanishes asymptotically with growing system size. Therefore, if U_{SIC}^v was to be finite, the localizing transformation \mathcal{M} could not be unity. Of course, there is always the possibility of a transformation between a Bloch and a localized base, but such a transformation does not necessarily lead to a valid solution for the localized orbitals in Eq. (2.4), as it may yield a total energy that is higher than the pure LDA solution. In some cases, such arbitrary localized orbitals even give a positive U_{SIC}^v , as

has been demonstrated by Norman³⁵ in the homogeneous electron gas. In fact, if U_{SIC}^v is not zero, the removal of a single electron from a valence band would lead to a localized hole in the density, which is not to be expected in standard semiconductors. We therefore assume that in the semiconductors we are going to consider, the localized and canonical valence states are identical and do not contribute to the self-interaction correction.

Earlier calculations have been performed with localized valence orbitals that have not been derived from a self-consistent minimization procedure. For example, very good band gaps have been obtained for Si in Ref. 36. However, such a method is, because of the arbitrariness in the choice of the localized orbitals, more comparable to semiempirical procedures. Their results show also that the wider gaps are obtained as a consequence of the shrinkage of the valence-band width as SIC causes more of the total energy to be transferred from the kinetic part to the potential part, in analogy to the situation in a Mott insulator. This makes it even more implausible that valence-band states in sp -bonded semiconductors possess a finite self-interaction correction.

III. SELF-INTERACTION-CORRECTED PSEUDOPOTENTIALS

If we assume *a priori* that there exists no self-interaction correction for the valence states, we can concentrate solely on the core states. The core states of a crystalline solid in a plane-wave calculation are customarily represented by pseudopotentials that are derived from all-electron self-consistent atomic calculations. The consistent construction of a pseudopotential requires that the atomic calculations are performed with the same method as the subsequent calculations in the solid. This is why, as mentioned in the Introduction, so many pseudopotentials are generated from atomic LDA calculations, even though it is well known that LDA in atoms has severe deficiencies. SIC-LDA offers a promising alternative since it produces good results in atoms^{6-8,21} and reduces for the valence states of semiconductors quite naturally to the simple LDA, as we know from the above considerations. Most importantly, the correct³ long-range $1/r$ behavior of the effective atomic potential is assured and the criterion that the highest occupied eigenvalue must equal the first ionization potential⁴ is much nearer fulfilled than in LDA.²¹ The consequence for the solid is eventually that the core density is better described than is done with LDA pseudopotentials.

We have developed a method to calculate the electronic structure of atoms with a fully relativistic self-interaction-corrected LDA,²¹ which enables us to generate pseudopotentials that can later reproduce the spin-orbit splitting in crystals. Let us briefly outline the basics of the atomic calculations and the pseudopotential construction.

We work in the spherical approximation and have to solve the following set of equations:

$$H_{\kappa} u_{m\kappa} = \epsilon_{m\kappa} u_{m\kappa}. \quad (3.1)$$

$u_{m\kappa}$ is the radial part, multiplied by r , of the canonical wave function, κ uniquely identifies the quantum numbers of orbital angular momentum l and total angular momentum j , and

$$\kappa = \begin{cases} -(l+1) & \text{if } j = l + \frac{1}{2} \\ l & \text{if } j = l - \frac{1}{2} \end{cases}. \quad (3.2)$$

The index m labels different eigenstates having the same quantum number κ and must not be confused with a magnetic quantum number. The radial functions $u_{m\kappa}$ are two-component spinor functions, comprising a major component G and a minor component F ,

$$u_{m\kappa}(r) = \begin{pmatrix} G_{m\kappa}(r) \\ F_{m\kappa}(r) \end{pmatrix}. \quad (3.3)$$

The Hamiltonian H_κ is given by

$$H_\kappa = H_D^\kappa - c^2 + v_{\text{eff}}^{\text{loc}}(r) + \sum_m (v^{\text{SIC}}[\rho_{m\kappa}] - \bar{v}^{\text{SIC}}) P_{m\kappa}, \quad (3.4)$$

$$H_D^\kappa = c \begin{pmatrix} c & -\frac{d}{dr} + \frac{\kappa}{r} \\ \frac{d}{dr} + \frac{\kappa}{r} & -c \end{pmatrix}. \quad (3.5)$$

In an atom of nuclear charge Z , the local part of the effective potential, including the local SIC contributions, is

$$v_{\text{eff}}^{\text{loc}}(r) = -\frac{Z}{r} + v_H([\rho]; r) + v_{\text{xc}}^{\text{RLDA}}(\rho(r)) + \bar{v}^{\text{SIC}}([\{\rho_i\}]; r), \quad (3.6)$$

where the ρ_i are spherically averaged orbital densities

$$\rho_i(r) = \frac{1}{r^2} [G_i^2(r) + F_i^2(r)]. \quad (3.7)$$

Here, all quantum numbers have been lumped together into i . The total density is consequently given by

$$\rho(r) = \sum_{i=1}^Z \rho_i(r). \quad (3.8)$$

$P_{m\kappa}$ is the projector onto the function

$$w_{m\kappa}(r) = \sum_{m'} \mathcal{M}_{m'm}^* u_{m'\kappa}(r), \quad (3.9)$$

which constitutes the radial part, multiplied by r , of the respective localized orbital. All other symbols are chosen as in Eqs. (2.4) and (2.5).

The eventual construction of the pseudopotentials is made more difficult by the existence of a nonlocal component in the effective potential and requires some additional thought. Our solution is to perform a two-step calculation. In the first step, a fully self-consistent relativistic SIC-LDA calculation is performed. In the second, we save the core density from the first calculation and

use it as input into a relativistic LDA calculation without SIC, in which only the valence density is determined self-consistently but the core density is kept fixed in its SIC-LDA shape. This method takes the picture of the "rigid core" seriously, which is at the heart of the pseudopotential method. One loses the strict orthogonality of core and valence states, but this orthogonality is never fully preserved as one moves the pseudopotential from the atomic environment to the solid. Since eventually the pseudopotentials will be used in a semiconductor, where the valence electrons see no self-interaction corrections, it is even more in the spirit of the problem to construct the pseudopotentials from LDA wave functions, while preserving the SIC effects on the core. From these latter LDA calculations, we construct the pseudopotentials in a standard manner, choosing the method of Hamann.³⁷

IV. RESULTS

The SIC potential is generally attractive and therefore causes the electrons to be closer concentrated around the atomic nucleus. This is demonstrated in Fig. 1, where the core density of the Ge atom from self-consistent calculations with and without SIC is compared. Consequently, the valence electrons in the LDA calculation see a better screened nucleus and become less tightly bound. This is mirrored in the pseudopotentials, which become less attractive when core self-interaction corrections are included (see Fig. 2).

Tables I and II give a compilation of the calculated eigenvalue spectra at selected Brillouin-zone points of Si, Ge, α -Sn, and GaAs. All calculations use relativistic

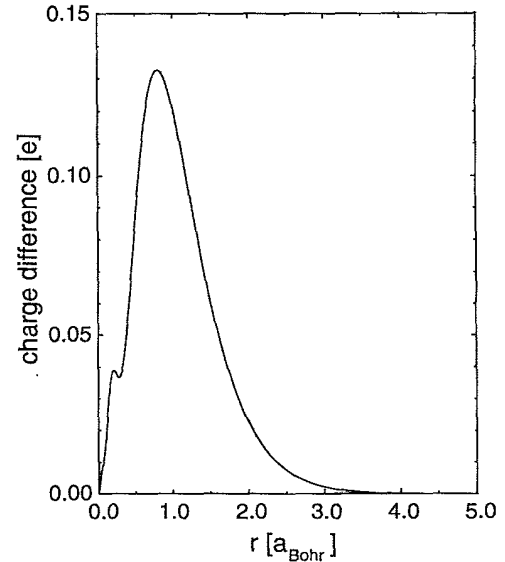


FIG. 1. Displayed is the integrated difference between the core charge densities of a free Ge atom with (SIC) and without (LDA) self-interaction corrections, that is the difference in core charge contained in a concentric sphere of radius r . SIC causes more of the charge to be concentrated near to the nucleus.

pseudopotentials^{38,39} and, with the exception of Si, spin-orbit coupling has been included in the crystal Hamiltonian. The Brillouin-zone integrations were done with ten special k points and the cutoff energies for the plane waves are 17 Ry in Si, 20 Ry in Ge, 16 Ry in Sn, and again 20 Ry in GaAs. The trend one generally observes is this: compared to pure LDA calculations, the lowest non-degenerate conduction band at Γ is moved upwards, as is the lowest conduction band at L , but to a lesser degree. The lowest conduction band at the X point moves very little, and in some cases the gap there shrinks slightly. Where the gaps increase, this means in all cases an improved agreement with experiment, as the LDA gaps are far too small.

The different size of the upward shift at Γ and L leads to an important effect in Ge, which is now described correctly as an indirect-gap semiconductor with conduction-band minimum at L , whereas LDA makes it appear a metal with a band overlap at Γ . The effect of SIC on the gaps in Ge is highlighted in Fig. 3. The gap error in Si cannot be corrected, since the fundamental gap is located close to the X point on the Δ axis. A fairly strong increase occurs in the energy of the $\Gamma_{2'}$ conduction band, but since this is not the lowest level at Γ and far above the fundamental gap, this is of little practi-

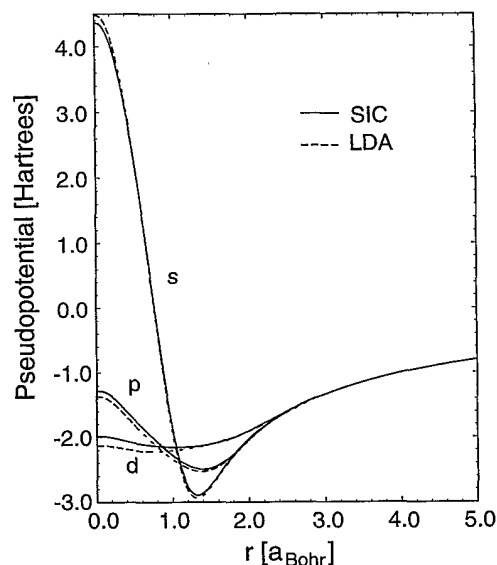


FIG. 2. Ge pseudopotentials for the s , p , and d angular momentum components (unscreened). The more tightly bound core charge of the SIC calculation screens the nucleus more effectively and results in a less attractive effective potential for the valence electrons, compared to simple LDA calculations.

TABLE I. Eigenvalue spectrum for Si and GaAs, calculated with ordinary LDA pseudopotentials and with pseudopotentials incorporating self-interaction corrections to the core (SIC). Stars on experimental values for GaAs indicate that they were measured without resolving the splitting of the valence-band top and have therefore been adjusted here by 0.13 eV.

Level	Si			Level	GaAs		
	SIC	LDA	Expt.		SIC	LDA	Expt.
$L_{2'v}$	-9.56	-9.56	-9.3 ± 0.4^a	L_{6v}	-11.14	-11.12	-11.37^{*a}
L_{1v}	-6.96	-6.95	-6.7 ± 0.2^a	L_{6v}	-6.75	-6.72	-6.83^{*a}
$L_{3'v}$	-1.16	-1.16	-1.2 ± 0.2^a	L_{6v}	-1.39	-1.35	-1.43^{*a}
L_{1c}	1.48	1.46	1.65 ± 0.01^a	$L_{4,5v}$	-1.15	-1.11	-1.43^{*a}
L_{3c}	3.26	3.34	4.15^b	L_{6c}	0.91	0.88	1.81^b
$L_{2'c}$	7.64	7.73		L_{6c}	4.46	4.52	5.32^{*c}
				$L_{4,5c}$	4.57	4.64	5.32^{*c}
				L_{6c}	7.59	7.82	8.47^{*c}
Γ_{1v}	-11.89	-11.88	-12.50 ± 0.6^a	Γ_{6v}	-12.75	-12.68	-13.08^{*a}
$\Gamma_{25'v}$	0.00	0.00	0.00	Γ_{7v}	-0.39	-0.39	-0.34^a
Γ_{15c}	2.56	2.59	3.05^c	Γ_{8v}	0.00	0.00	0.00
$\Gamma_{2'c}$	3.41	3.26	4.1^c	Γ_{6c}	0.57	0.46	1.52^a
				Γ_{7c}	3.49	3.54	4.488^a
				Γ_{8c}	3.70	3.76	4.659^a
X_{1v}	-7.76	-7.77		X_{6v}	-10.44	-10.44	-10.88^{*a}
X_{4v}	-2.81	-2.81	-2.9^d	X_{6v}	-6.86	-6.90	-6.83^{*a}
X_{1c}	0.54	0.62	1.25^c	X_{6v}	-2.86	-2.78	-2.93^{*a}
X_{4c}	10.18	10.11		X_{7v}	-2.77	-2.69	-2.93^{*a}
				X_{6c}	1.18	1.27	1.98^a
				X_{7c}	1.45	1.49	$2.38^{a,c}$
				X_{6c}	10.14	10.13	
				X_{7c}	10.22	10.20	

^aReference 46.

^bReference 47.

^cReference 48.

^dReference 49.

TABLE II. Like Table I, now for Ge and Sn. Again, stars on experimental values for Ge indicate that they were measured without resolving the splitting of the valence-band top and have therefore been adjusted here by 0.1 eV.

Level	Ge			Sn		
	SIC	LDA	Expt.	SIC	LDA	Expt.
L_{6v}^-	-10.70	-10.72	$-10.7 \pm 0.5^{*a}$	-9.28	-9.24	
L_{6v}^+	-7.70	-7.65	$-7.8 \pm 0.2^{*a}$	-6.62	-6.58	
L_{6v}^-	-1.62	-1.57	$-1.5 \pm 0.3^{*a}$	-1.69	-1.66	
$L_{4,5v}$	-1.41	-1.36	$-1.5 \pm 0.3^{*a}$	-1.21	-1.18	
L_{6c}^-	0.10	0.03	0.744 ± 0.1^a	-0.29	-0.30	$0.09 - 0.115^a$
L_{6c}^+	3.57	3.59	4.3^{*b}	2.81	2.86	
$L_{4,5c}$	3.70	3.72	4.3^{*b}	3.08	3.14	
L_{6c}^-	6.98	7.13	7.7^{*b}	5.43	5.59	
Γ_{6v}^+	-12.76	-12.75	$-13.0 \pm 0.3^{*c}$	-10.89	-10.80	
Γ_{7v}^+	-0.34	-0.34	-0.296^a	-0.75	-0.75	-0.80^a
Γ_{8v}^+	0.00	0.00	0.00	0.00	0.00	0.00
Γ_{7c}^-	0.12	-0.10	0.898^a	-0.80	-0.90	-0.634 to -0.413^a
Γ_{6c}^-	2.31	2.33	$3.006^{a,d}$	1.55	1.58	1.98^a
Γ_{8c}^-	2.56	2.58	$3.206^{a,d}$	2.08	2.12	$2.25 - 2.48^a$
X_5	-8.93	-8.95	$-9.4 \pm 0.3^{*c}$	-7.82	-7.79	
X_5	-3.20	-3.10	$-3.25 \pm 0.20^{*a}$	-2.80	-2.75	
X_5	0.53	0.56	$1.2 \pm 0.20^{*a}$	0.45	0.51	
X_5	9.53	9.43		7.52	7.50	

^aReference 46.

^bReference 48.

^cReference 50.

^dReference 51.

cal relevance. In GaAs, the lowest conduction bands at Γ and L move in the right direction, reducing the LDA error as they do in α -Sn, but this is not enough to eliminate the spurious band overlap at L . In all materials, the spin-orbit splitting of the valence-band edge is described in fair agreement with experiments, independent of any self-interaction correction.

The clue to these trends can be taken from simple estimates. The tighter core screens the nucleus more effectively and therefore causes the atomic valence states, when calculated without self-interaction corrections, to be weaker bound. The s electrons experience more of this effect than the outermost p electrons, and as a con-

sequence the level distance between the s and p valence electrons shrinks, compared to a normal LDA calculation (see Table III). Chadi and Cohen⁴⁰ have given the following tight-binding formulas:

$$E(\Gamma_{2'}^c) - E(\Gamma_{25'}^v) = \epsilon_s - \epsilon_p + V_1, \quad (4.1)$$

$$E(X_1^c) - E(\Gamma_{25'}^v) = \frac{1}{2}(\epsilon_s - \epsilon_p) + \frac{1}{2}\sqrt{(\epsilon_s - \epsilon_p)^2 + 4V_2^2}, \quad (4.2)$$

$$E(L_1^c) - E(\Gamma_{25'}^v) = \frac{1}{2}(\epsilon_s - \epsilon_p) - \frac{1}{2}\sqrt{(\epsilon_s - \epsilon_p + V_3)^2 + 12V_2^2}, \quad (4.3)$$

$$E(L_{15c}^c) - E(\Gamma_{25'}^v) = V_4, \quad (4.4)$$

valid for diamond structure materials like Si, Ge, and α -

TABLE III. Eigenvalue differences $\epsilon_s - \epsilon_p$ (in eV) of the highest s and p levels in the free atom, calculated in relativistic LDA, or in relativistic LDA for the valence electrons and a fixed core density from relativistic all-electron SIC-LDA calculations. In the latter case, the level distances generally decrease.

Element	LDA	LDA + SIC core
Si	-6.72	-6.68
Ge	-7.90	-7.77
Sn	-6.93	-6.83
Ga	-6.43	-6.28
As	-9.35	-9.25

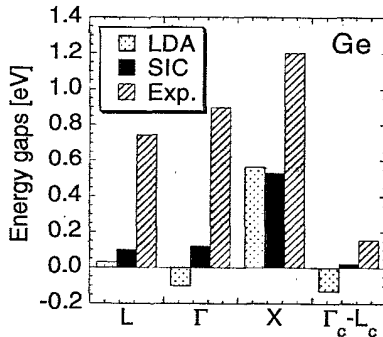


FIG. 3. Selected energy gaps in Ge. Self-interaction corrections open up the gaps at L and Γ , but by different amounts, so that the fundamental gap becomes indirect. The gap at X shrinks by a very small amount.

TABLE IV. Theoretical and experimental lattice constants a (in Å) and bulk moduli B (in GPa). Experimental values from Ref. 46.

	a_{LDA}	a_{SIC}	a_{exp}	B_{LDA}	B_{SIC}	B_{expt}
Si	5.39	5.44	5.43	97.9	96.3	97.8
Ge	5.58	5.70	5.66	77.8	73.7	75.8
Sn	6.44	6.53	6.49	49.8	46.1	53.0
GaAs	5.57	5.70	5.66	79.0	72.5	76.9

Sn, to predict the band structure in the solid from the atomic s and p valence levels. The V_i denote compounded overlap matrix elements that depend themselves only in higher order on the atomic eigenvalues and they can be assumed constant in a rough estimate. One sees now immediately that as a consequence of the shrinkage of the atomic $s - p$ level distance the gap between Γ_2^c and Γ_{25}^v will open up. The indirect gap $L_{1c} - \Gamma_{25}^v$ will also widen, but to a lesser degree, since V_3 is positive and of the same order of magnitude as $\epsilon_s - \epsilon_p$ in all cases considered.⁴⁰ The gap between X_1^c and Γ_{25}^v will vary least, as there are positive and negative contributions almost canceling each other. For a weakly polar zinc-blende material like GaAs, these qualitative estimates remain valid.

Beside the eigenvalue gaps, there is also a clear effect of SIC on the lattice constants. Table IV lists the lattice constants calculated with and without SIC. Generally SIC widens the lattice constants, as the better screened nuclei cause more charge to be concentrated in the interstitial regions. Even though LDA describes the lattice constants fairly well, the relative error is once more reduced by SIC. At the same time, the bulk modulus, calculated from a fit of the total energy to the Murnaghan equation of state,⁴¹ is generally somewhat reduced compared to LDA and below the experimental value.

Some improvement of standard LDA pseudopotential results is possible, if nonlinear core corrections⁴² are taken into account. This leads unfortunately to even smaller gaps but improvements in the lattice constants, as has been shown, for example, by Qteish and Needs⁴³ for several semiconductors.

In our current method we retain a nonlinear exchange-correlation potential. One might expect the effects of any nonlinear core correction to be less important than in pure LDA since the overlap between the SIC-LDA core and valence density is smaller. On applying these partial core corrections⁴² (PCC) to the present SIC-LDA scheme we see effects that are small indeed but exactly analogous to normal LDA: the gaps shrink, the valence-band width widens, and the lattice constants increase slightly.

We give here two examples for the effects of PCC. Using LDA+PCC and SIC+PCC, respectively, for the lattice constants, we find 5.40 and 5.44 Å in Si and 5.61 and 5.72 Å in Ge. The direct gap is obtained by the two methods as 2.57 and 2.54 eV in Si and -0.16 and 0.05 eV in Ge, respectively, and Ge remains just indirect. Thus, all the qualitative statements made earlier remain true as far as the gaps are concerned, but the SIC-LDA plus PCC results for the lattice constants do not agree better or worse with experiment than LDA plus PCC calculations in general.

In passing we note that the question of an improved description of the core has been tackled with a quite different method by Bylander and Kleinman,⁴⁴ who constructed pseudopotentials where in the atomic calculations the core electrons contributed to a Hartree-Fock potential and the valence electron to an LDA exchange-correlation potential. They thus could also more accurately describe the core density and observed qualitatively similar effects as we do here but achieved quantitatively even better agreement between theory and experiment.^{44,45}

V. SUMMARY

We have generated pseudopotentials from fully self-consistent relativistic self-interaction-corrected LDA calculations in atoms and thus achieved important improvements in the description of the core density. We could demonstrate that, compared to standard LDA pseudopotentials, these pseudopotentials lead to a widening of the eigenvalue gaps at the Γ and L point of the Brillouin zone in the materials considered. Also the distance between the lowest conduction band at Γ and L is brought better into line with experiment. We can trace these improvements to a tighter screening of the atomic nuclei, an effect that also leads to an increase in the lattice constants. In no case can the gap problem of LDA be remedied in its full extent, though. The lattice constants can be calculated with a precision that is superior to simple LDA and comparable to the LDA with nonlinear core corrections.

ACKNOWLEDGMENTS

This work greatly benefited from many discussions with and practical help from Dr. J. A. Majewski and Dr. A. Seidl. Financial support was provided by Deutsche Forschungsgemeinschaft (SFB 348), Bayerischer Forschungverbund FOROPTO, and Deutsche Volkswagenstiftung, Projekt I/67998.

* Present address: Cavendish Laboratory, TCM Group, Madingley Road, Cambridge, CB3 0HE, UK.

¹ P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

² W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

³ O. Gunnarson, M. Jonson, and B. I. Lundqvist, Phys. Rev. B **20**, 3136 (1979).

⁴ J. P. Perdew, R. G. Parr, M. Levy, and J. Jose L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982).

⁵ U. von Barth and L. Hedin, J. Phys. C **5**, 1629 (1972).

⁶ J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

⁷ J. G. Harrison, R. A. Heaton, and C. C. Lin, J. Phys. B **16**, 2079 (1983).

- ⁸ J. G. Harrison, *J. Chem. Phys.* **78**, 4562 (1983).
- ⁹ M. R. Pederson, R. A. Heaton, and C. C. Lin, *J. Chem. Phys.* **80**, 1972 (1984).
- ¹⁰ M. R. Pederson, R. A. Heaton, and C. C. Lin, *J. Chem. Phys.* **82**, 2688 (1985).
- ¹¹ R. A. Heaton, J. G. Harrison, and C. C. Lin, *Solid State Commun.* **41**, 827 (1982).
- ¹² R. A. Heaton, J. G. Harrison, and C. C. Lin, *Phys. Rev. B* **28**, 5992 (1983).
- ¹³ R. A. Heaton and C. C. Lin, *J. Phys. C* **17**, 1853 (1984).
- ¹⁴ S. C. Erwin and C. C. Lin, *J. Phys. C* **21**, 4285 (1988).
- ¹⁵ A. Svane, *Phys. Rev. Lett.* **68**, 1900 (1992).
- ¹⁶ A. Svane, *Phys. Rev. Lett.* **72**, 1248 (1994).
- ¹⁷ Z. Szotek, W. M. Temmerman, and H. Winter, *Phys. Rev. B* **47**, 4029 (1993).
- ¹⁸ W. M. Temmerman, Z. Szotek, and H. Winter, *Phys. Rev. B* **47**, 1184 (1993).
- ¹⁹ W. M. Temmerman, Z. Szotek, and H. Winter, *Phys. Rev. B* **47**, 11533 (1993).
- ²⁰ Z. Szotek, W. M. Temmerman, and H. Winter, *Phys. Rev. Lett.* **72**, 1244 (1994).
- ²¹ M. M. Rieger and P. Vogl, *Phys. Rev. A* **52**, 282 (1995).
- ²² G. B. Bachelet, D. R. Hamann, and M. Schlüter, *Phys. Rev. B* **26**, 4199 (1982).
- ²³ G. P. Kerker, *J. Phys. C* **13**, 189 (1980).
- ²⁴ N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 1992 (1991).
- ²⁵ N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 8861 (1991).
- ²⁶ A. Zunger, *Phys. Rev. B* **22**, 649 (1979).
- ²⁷ E. L. Shirley, R. M. Martin, G. B. Bachelet, and D. M. Ceperley, *Phys. Rev. B* **42**, 5057 (1990).
- ²⁸ L. J. Sham, *Phys. Rev.* **150**, 720 (1966).
- ²⁹ J. P. Perdew and M. Levy, *Phys. Rev. Lett.* **51**, 1884 (1983).
- ³⁰ L. J. Sham and M. Schlüter, *Phys. Rev. Lett.* **51**, 1888 (1983).
- ³¹ R. W. Godby, M. Schlüter, and L. J. Sham, *Phys. Rev. B* **36**, 6497 (1987).
- ³² R. W. Godby, M. Schlüter, and L. J. Sham, *Phys. Rev. B* **37**, 10159 (1988).
- ³³ O. Gunnarson and K. Schönhammer, *Phys. Rev. Lett.* **56**, 1968 (1986).
- ³⁴ K. Schönhammer and O. Gunnarson, *J. Phys. C* **20**, 3675 (1987).
- ³⁵ M. R. Norman, *Phys. Rev. B* **28**, 3585 (1983).
- ³⁶ M. Ikeda, K. Terakura, and T. Oguchi, *Phys. Rev. B* **48**, 1571 (1993).
- ³⁷ D. R. Hamann, *Phys. Rev. B* **40**, 2980 (1989).
- ³⁸ L. Kleinman, *Phys. Rev. B* **21**, 2630 (1980).
- ³⁹ G. B. Bachelet and M. Schlüter, *Phys. Rev. B* **25**, 2103 (1982).
- ⁴⁰ D. J. Chadi and M. L. Cohen, *Phys. Status Solidi B* **68**, 405 (1975).
- ⁴¹ F. D. Murnaghan, *Proc. Natl. Acad. Sci.* **30**, 244 (1944).
- ⁴² S. G. Louie, S. Froyen, and M. L. Cohen, *Phys. Rev. B* **26**, 1738 (1982).
- ⁴³ A. Qteish and R. J. Needs, *Phys. Rev. B* **43**, 4229 (1991).
- ⁴⁴ D. M. Bylander and L. Kleinman, *Phys. Rev. B* **43**, 12070 (1991).
- ⁴⁵ Y. M. Gu, D. M. Bylander, and L. Kleinman, *Phys. Rev. B* **50**, 2227 (1994).
- ⁴⁶ *Numerical Data and Functional Relationships in Science and Technology*, edited by K.-H. Hellwege and O. Madelung, Landolt-Börnstein, New Series, Group III, Vol. 17, Pt. a (Springer, Berlin, 1982).
- ⁴⁷ F. J. Himpsel, P. Heimann, and D. E. Eastman, *Phys. Rev. B* **24**, 2003 (1981).
- ⁴⁸ J. E. Ortega and F. J. Himpsel, *Phys. Rev. B* **47**, 2130 (1993).
- ⁴⁹ W. E. Spicer and R. C. Eden, in *Proceedings of the 9th International Conference on the Physics of Semiconductors*, edited by S. M. Ryvkin (Nauka, Moscow, 1968), Vol. 1, p. 68.
- ⁵⁰ A. L. Wachs *et al.*, *Phys. Rev. B* **32**, 2326 (1985).
- ⁵¹ D. E. Aspnes, *Phys. Rev. B* **12**, 2297 (1975).