PHYSICAL REVIEW B VOLUME 36, NUMBER 6 15 AUGUST 1987-II

Nonlocal Kohn-Sham exchange corrections to Si band gaps and binding energy

Yu-Tang Shen, D. M. Bylander, and Leonard Kleinman
Department of Physics, University of Texas, Austin, Texas 78712

(Received 6 February 1987)

We compare the binding energy and energy band gaps of Si calculated with the local-density approximation to the Kohn-Sham exchange potential and with the correction to that potential which is exact to first order in $\rho(\mathbf{K})/\rho_0$. Although the correction is not small, running between 1 eV and -3.5 eV in the unit cell, its effect on both the band gaps and binding energy is very small. Thus we conclude that the discrepancy between calculated energy band gaps and experiment must be almost entirely due to the discontinuity in the zeroth Fourier component of the exchange-correlation potential.

A few years ago Perdew and Levy[1] and Sham and Schlüter[2] demonstrated that there is a finite correction to the band gap of a semiconductor obtained from the exact eigenvalues of the Kohn-Sham[3] equation due to a discontinuity Δ in the zeroth Fourier transform of the exchange-correlation potential. Kohn-Sham calculations in the local-density approximation (LDA) are known to result in large energy-gap errors. For both Si (this paper) and Ge (Ref. 4) the calculated gap is about 0.75 eV too small, which results in Ge becoming a zero-gap semiconductor.[4] One-dimensional model calculations[5,6] indicated that Δ is large enough to account for the gap error. Because for kinetic energy one and three dimensions are essentially equivalent and because everything could be calculated analytically, we[7] found it pedagogically useful to calculate the discontinuity in the kinetic potential $\delta T_s/\delta n$. In the same paper we speculated that the discontinuity in the exchange-correlation potential in real crystals might be smaller than originally thought and that a large part of the gap error might arise from using the LDA. Because the Hartree-Fock gap in Si is about three times larger than the experimental gap,[8] it is obvious that the discontinuity in the Kohn-Sham (KS) exchange potential is very large and that the discontinuity in the KS correlation potential cancels a large fraction of it. The only question is, is that fraction closer to 95% or to 75%? There does not appear to be any direct way to calculate the exchange-correlation potential discontinuity in a real three-dimensional semiconductor. Therefore the only way (other than from one-dimensional models) to estimate Δ is to systematically improve the KS potential beyond the LDA until one is confident that the only remaining gap error is the neglect of Δ. There have been many attempts to improve upon the LDA. Levine and Louie[9] constructed the dielectric function for a free-electron gas with an energy gap whence, in the manner of Singwi,[10] they constructed a V_{XC}. This resulted in an increase in the exchange part[11] of V_{XC} with increasing gap which we consider to be unphysical within their model. Since exchange is a one-electron property, the introduction of a gap while keeping single-plane-wave wave functions should have no effect. Also a gap *per se* should reduce the correlation potential. Thus the small increases they obtain in Si energy gaps, we believe, are fortuitous. Hybertsen and Louie[12] performed weighted-density-approximation[13] (WDA) calculations. The WDA uses an *ad hoc* exchange-correlation (XC) hole which is constrained to yield the desired XC potential and contain one electron in the homogeneous limit only. Still, the approximation seems physically reasonable and one might suspect that the 0.15-eV increase they obtained in the Si indirect gap is about as large as can be obtained with nonlocal corrections to the LDA. A previous WDA calculation by Kerker[14] had resulted in much larger energy-gap corrections, but he had made an unjustified approximation[15] in $V_{\mathrm{XC}}=\delta E_{\mathrm{XC}}/\delta n$ which made his V_{XC} equivalent to a Slater potential rather than a KS potential. It has been known for a long time[16] that the Slater potential yields better energy gaps than the KS in the LDA. von Barth and Car[17] found no improvement in Si energy gaps using the Langreth-Mehl[18] gradient correction to the KS XC potential. There have been many calculations of energy-dependent self-energies culminating in the most impressive work of Hybertsen and Louie,[19] all of whose calculated Si energy gaps were in excellent agreement with experiment. These self-energy calculations are, however, beyond the realm of density-functional theory.

It is best in density-functional theory to define the exchange energy to be the Fock energy of a ground-state configuration consisting of an antisymmetrized product of the N lowest Kohn-Sham eigenfunctions. This energy is of order e^2/r_s and contains no higher-order corrections.[20] Correlation energy may be obtained in principle by including the interaction with all other configurations made up of KS eigenfunctions. The correlation energy contains all terms of order $(e^2r_s)^n e^2/r_s$ with $n \geq 1$. All the corrections to the LDA we have discussed contain correlation terms of order e^2/r_s. These (presumed) errors arise from either constructing an *ad hoc* XC hole or calculating one but making certain approximations such as static rather than dynamic screening. It has been suggested[21] that singularities in the correlation energy could, in fact, cause it to have terms with the dimensionality of exchange. We do not believe this can ever be the case; for semiconductors, where there are no singularities, this is certainly not the case.[22] For Si with $r_s=2$ correlation is only 14% of exchange in the LDA. Therefore it was our hope that if corrections to the LDA exchange were done accurately, and corrections to the LDA correlation ignored, large corrections to the band gaps might be obtained. In the remainder of this Brief Report we describe such a calculation of both the energy bands and binding energy of Si (at the equilibrium lattice constant, $a = 10.261$ bohr), and explain why even though the correction to V_x is large, it results in negligible corrections to both the energy gaps and binding energy.

We constructed a scalar relativistic nonlocal norm-conserving pseudopotential[23] for Si, used the Wigner correlation potential, and expanded in a mixed basis set consisting of all plane waves with $k^2 < 32.3(2\pi/a)^2$ plus two *s*, one *p*, and one *d* Gaussians. At Γ we compared with a presumably converged calculation which used a 965-plane-wave expansion and found $\Gamma_{25'}$ and Γ_1 unconverged by only 2.9 and 0.3 meV, respectively. In our self-consistent calculation with LDA exchange we found the indirect gap to be 0.11 eV smaller than similar calculations by others.[19] This is due at least in part to relativistic effects.

We[24,25] have derived a formula for the exchange energy of an electron gas responding to a weak external potential $\sum_{\mathbf{G}} V_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}$, which is *exact* to second order in $\rho(\mathbf{G})/\rho_0$,

$$E_x = -\frac{3}{2}\left[\frac{3}{\pi}\right]^{1/3}\left[\rho_0^{4/3} + \rho_0^{-2/3}\sum_{\mathbf{G}}{}' \rho(\mathbf{G})\rho^*(\mathbf{G})F(G/k_F)\right], \tag{1}$$

where $F(G/k_F)$ is a universal curve plotted in Fig. 1 of Ref. 25. This results in an exchange potential

$$V_x = -3\left[\frac{3}{\pi}\right]^{1/3} \rho_0^{-2/3} \sum_{\mathbf{G}} \rho^*(\mathbf{G}) F(G/k_F) e^{-i\mathbf{G}\cdot\mathbf{r}} . \tag{2}$$

Here E_x and V_x are in rydbergs when ρ is in e^-/bohr^3. We noted that $F_{\text{LDA}} = F(G=0) = \frac{2}{9}$ and suggested that[25]

$$\delta V_x = -3\left[\frac{3}{\pi}\right]^{1/3} \rho_0^{-2/3} \sum_{\mathbf{G}} \rho^*(\mathbf{G})[F(G/k_F) - \tfrac{2}{9}] e^{-i\mathbf{G}\cdot\mathbf{r}} \tag{3}$$

be used as a correction to the LDA KS exchange potential. Here $\rho(\mathbf{G})$ is the **G**th Fourier component of the valence-electron pseudocharge density. Our LDA valence V_{XC} was taken[26] to be the difference between $V_{\text{XC}}(\rho_{\text{total}})$ and $V_{\text{xv}}(\rho_{\text{core}})$ and thus we were unable to compare the use of (2) directly with the use of (3) as a correction in order to test the convergence of the expansion in powers of $\rho(\mathbf{G})/\rho_0$. A self-consistent calculation was performed

TABLE I. $[F(G/k_F) - \frac{2}{9}]$ and $\rho(\mathbf{G})$ in 10^{-3} e^-/bohr^3 for several Si reciprocal lattice vectors $\mathbf{G} = (2\pi/a)(i,j,k)$. The origin of coordinates is at the point midway between bonding atoms ensuring real $\rho(\mathbf{G})$.

(i,j,k)	$F(G/k_F) - 2/9$	$\rho(\mathbf{G})$
(1,1,1)	0.0462	−6.5619
(2,2,0)	0.1920	0.2421
(3,1,1)	−0.0594	1.2323
(2,2,2)	−0.0972	1.1911
(4,0,0)	−0.1515	1.3329
(3,3,1)	−0.1681	−0.3784
(4,2,2)	−0.1830	−0.5830
(3,3,3)	−0.1886	−0.4570
(5,1,1)	−0.1886	−0.3283
(4,4,0)	−0.1950	−0.2441
(5,3,1)	−0.1977	−0.1545

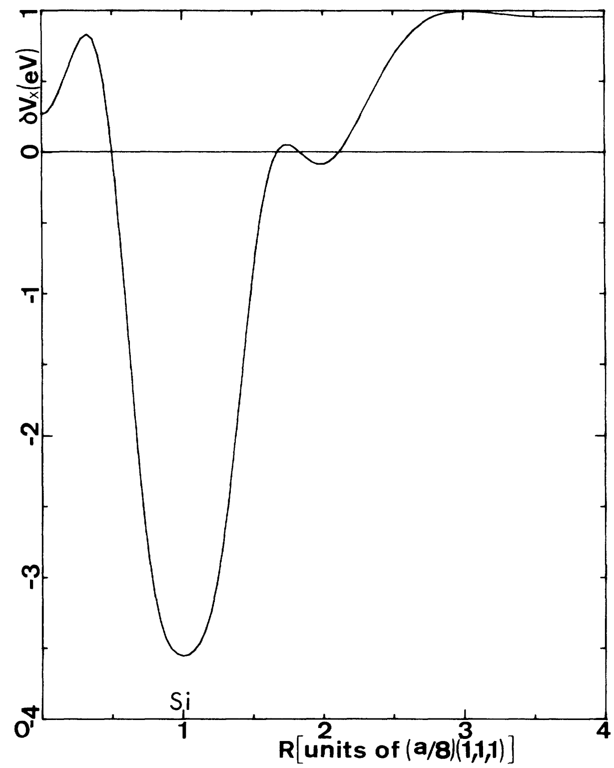

FIG. 1. Plot of δV_x (in eV) along the [111] axis of the crystal (in units of $\sqrt{3}a/8$). The origin is at the center of the covalent bond and the position of the Si atom is marked.

TABLE II. Energy of several states relative to the top of the valence-band $\Gamma_{25'}$ state (in eV) without and with the exchange correction δV_x. The experimental values are taken from Ref. 19.

State	LDA	δV_x	Expt.
Γ_1	−11.992	−11.994	−12.5±0.6
Γ_{15}	2.499	2.441	3.4
$\Gamma_{2'}$	3.207	3.154	4.2
X_{1v}	−7.836	−7.862	
X_{4v}	−2.880	−2.894	−2.9,−3.3±0.2
X_{1c}	0.546	0.655	1.3
X_{4c}	9.967	9.899	
$L_{2'}$	−9.640	−9.670	−9.3±0.4
L_{1v}	−7.035	−7.014	−6.7±0.2
$L_{3'}$	−1.205	−1.225	−1.2±0.2, −1.5
L_{1c}	1.405	1.432	2.1,2.4±0.15
L_3	3.257	3.239	4.15±0.1
Δ_1(gap)	0.412	0.514	1.17

TABLE III. Contributions to the total valence-electron energy of Si (in Ry per atom) without and with the exchange correction.

	LDA	δV_x
$\sum_{n\mathbf{k}}\varepsilon_{n\mathbf{k}}-\sum_{\mathbf{G}}V(\mathbf{G})\rho(\mathbf{G})$	2.573 470	2.572 891
$\frac{1}{2}8\pi\Omega\sum_{\mathbf{G}}\rho(\mathbf{G})^2/G^2$	0.535 883	0.545 746
$E_{\mathrm{XC}}^{\mathrm{LDA}}(\rho_{\mathrm{total}})-E_{\mathrm{XC}}^{\mathrm{LDA}}(\rho_{\mathrm{core}})$	−2.704 793	−2.713 471
δE_x		−0.000 060
E_{Ewald}	−8.399 646	−8.399 646
E_{total}	−7.995 086	−7.994 540

TABLE IV. Nonrelativistic Hartree-Fock and all-electron LDA valence-electron binding energies (in Ry) obtained from the difference between Si atomic and Si^{4+} ionic binding energies compared with the scalar relativistic LDA pseudopotential binding energy and experiment.

	Atom	Ion	Valence
HF	577.708 8	570.362 0	7.346 8
LDA (all)	575.755 40	568.137 95	7.617 45
LDA (pseudo)			7.622 72
Expt.			7.576 6

with the additional potential δV_x and in Fig. 1 the self-consistent δV_x is plotted along the [111] axis; it is seen to range between +1 and −3.5 eV. In Table I are listed the values of $F(\mathbf{G}/k_F)-\frac{2}{9}$ and $\rho(\mathbf{G})$ which contribute to δV_x for the first few sets of plane waves. Plane waves out to (14,6,4) were used in the actual calculation.

In Table II energy levels at Γ, X, and L are listed relative to $\Gamma_{25'}$. The inclusion of δV_x is seen to improve the indirect gap by only 0.10 eV and to worsen the direct gap by 0.06 eV. It is at first surprising that the large δV_x of Fig. 1 yields such small corrections to the energy gaps. What seems to be happening is that the higher-wave-number corrections are canceling the $\mathbf{G}=(2\pi/a)(1,1,1)$ correction. For example, the $\Gamma_{25'}$ matrix element between (111) and (200) symmetrized combinations of plane waves is $2(V_{111}-V_{311})$. This matrix element is large because V_{111} is negative and V_{311} positive and is responsible for $\Gamma_{25'}$ lying below $\Gamma_{2'}$ and Γ_{15}. From Eq. (3) and Table I it is found that $\delta V_x(111)=-0.138$ eV and $\delta V_x(311)=-0.033$ eV and thus they tend to cancel in this matrix element. The total drop in $\Gamma_{25'}$ due to δV_x is 0.066 eV, which is even less than would be expected from this matrix element due to canceling contributions from other matrix elements.

In Table III are listed contributions to the total energy of Si with and without exchange corrections calculated in the usual variational manner.[27] δE_x is given by the second term of Eq. (1) with F replaced by $F-\frac{2}{9}$. We note that because the $\rho(\mathbf{G})$ in Table I are of order 10^{-3} e^-/bohr^3 and δE_x is quadratic in the $\rho(\mathbf{G})$, its contribution to E_{total} is an order of magnitude smaller than (and of opposite sign to) the second-order effects of δV_x. (Note that the first term in the table, the sum of one-electron eigenvalues minus the potential due to the valence electrons times their charge density, contains no first-order contributions from the valence-electron potential.) The net effect of the exchange corrections to the LDA is to reduce the binding energy ($-E_{\mathrm{total}}$) by a negligible 7.4 meV. We have not calculated the cohesive energy in Table III because our exchange correction is not readily applicable to atoms. Within the LDA if we add the zero-point energy of[28] 0.005 Ry to E_{total} and subtract our calculated spin-polarized atomic energy of −7.622 716 Ry we obtain a cohesive energy of 4.999 eV to compare with the experimental value of 4.63 eV and the nonrelativistic 4.83 eV calculated by Yin and Cohen.[28]

Thus we have demonstrated that a rather large correction to the LDA exchange potential has a very small effect on Si energy gaps and a negligible effect on its binding energy. In closing, we would like to suggest that core-valence exchange corrections which we have ignored might be more significant than the valence-valence corrections considered here. In Table IV we compare nonrelativistic atomic all-electron Hartree-Fock[29] and LDA valence binding energies with our relativistic LDA pseudopotential result and the experimental Si fourfold ionization energy.[30] Note that the HF exchange makes a larger contribution to the binding of both the Si atom and Si^{4+} ion than the LDA exchange plus correlation but that the LDA valence electrons are more bound by 3.68 eV. This may be attributed to better screening by the more compact HF cores and to a larger core orthogonalization kinetic energy of the HF valence electrons. Thus if a pseudopotential were constructed which was based on an HF core, the effects on the Si binding energy (and perhaps cohesive energy also) would be non-negligible. The effects on the band gaps are much harder to estimate but we would guess that they are small since the HF and LDA atomic $3s$-$3p$ splittings are nearly identical.

This work was supported by the Robert A. Welch Foundation (Houston, Texas) and the National Science Foundation under Grant No. DMR-8412408.

---

[1] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).

[2] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).

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

[4] G. B. Bachelet and N. E. Christensen, Phys. Rev. B **31**, 879 (1985). The spin-orbit splitting must be added to this calculation to obtain a zero-gap semiconductor.

[5] L. J. Sham and M. Schlüter, Phys. Rev. B **32**, 3883 (1985).

[6] M. Lannoo, M. Schlüter and L. J. Sham, Phys. Rev. B **32**, 3890 (1985).

[7] L. Kleinman, Phys. Rev. B **33**, 7299 (1986).

[8] J. C. Phillips and L. Kleinman, Phys. Rev. **128**, 2098 (1962).

[9] Z. H. Levine and S. G. Louie, Phys. Rev. B **25**, 6310 (1982).

[10] K. S. Singwi, A. Sjölander, M. P. Tosi, and R. H. Land, Phys. Rev. B **1**, 1044 (1970).

[11] Expanding their V_{xc} into terms proportional to $r_s^{-1}, r_s^0, r_s, r_s^2, \ldots$ we assume that the leading term is exchange and the rest correlation.

[12] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **30**, 5777 (1984).

[13] O. Gunnarsson and R. O. Jones, Phys. Scr. **21**, 394 (1980).

[14] G. P. Kerker, Phys. Rev. B **24**, 3468 (1981).

[15] C. S. Wang and W. E. Pickett, Phys. Rev. Lett. **51**, 597 (1983).

[16] D. J. Stukel and R. N. Euwema, Phys. Rev. B **1**, 1635 (1970).

[17] U. von Barth, in *The Electronic Structure of Complex Systems,* Vol. 113 of *NATO ASI Series B,* edited by P. Phariseau and W. M. Temmerman (Plenum, New York, 1984), p. 67.

[18] D. C. Langreth and M. J. Mehl, Phys. Rev. B **28**, 1809 (1983).

[19] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[20] It is by now well known that if Hartree-Fock eigenfunctions are used, the same result is obtained to first order in e^2 but higher-order terms in e^2 are also obtained which are infinite. The infinities are canceled by infinities in the correlation energy. See Refs. 25 and 24 and references therein.

[21] John Perdew (private communcation) points out that a product of screened Coulomb interactions gives rise to these e^2/r_s correlation terms. The screening, of course, arises from summing an infinite set of singular diagrams. In the LDA the corresponding term is a product of screened and unscreened Coulomb interactions which gives rise to a $\ln r_s$ term. The original explanation for these e^2/r_s terms was given by S. K. Ma and K. A. Brueckner, Phys. Rev. **165**, 18 (1968), but does not stand up under scrutiny.

[22] Not only are there no vanishing energy denominators in a semiconductor but there are also no q^{-2} matrix-element singularities because of the orthogonality of states in different bands.

[23] D. M. Bylander and L. Kleinman, Phys. Rev. B **29**, 1534 (1984), Appendix A.

[24] L. Kleinman, Phys. Rev. B **30**, 2223 (1984).

[25] P. R. Antoniewicz and L. Kleinman, Phys. Rev. B **31**, 6779 (1985).

[26] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[27] D. M. Bylander and L. Kleinman, Phys. Rev. B **34**, 5280 (1986).

[28] Y. T. Yin and M. L. Cohen, Phys. Rev. B **26**, 5668 (1982).

[29] S. Fraga, J. Karwowski, and K. M. S. Saxena, *Handbook of Atomic Data* (Elsevier-Scientific, Amsterdam, 1976), p. 141.

[30] C. E. Moore, *Atomic Energy Levels,* Natl. Bur. Stand. (U.S.) Circ. No. 467 (U.S. GPO, Washington, D.C., 1949), Vol. I.