Effect of semicore orbitals on the electronic band gaps of Si, Ge, and GaAs within the GW approximation

Murilo L. Tiago, Sohrab Ismail-Beigi,* and Steven G. Louie
Department of Physics, University of California at Berkeley, Berkeley, California 94720-7300, USA
and Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
(Received 8 July 2003; published 19 March 2004)

We study the effect of semicore states on the self-energy corrections and electronic energy gaps of Si, Ge, and GaAs. Self-energy effects are computed within the GW approach, and electronic states are expanded in a plane-wave basis. For these materials, we generate *ab initio* pseudopotentials treating as valence states the outermost two shells of atomic orbitals, rather than only the outermost valence shell as in traditional pseudopotential calculations. The resulting direct and indirect energy gaps are compared with experimental measurements and with previous calculations based on pseudopotential and ''all-electron'' approaches. Our results show that, contrary to recent claims, self-energy effects due to semicore states on the band gaps can be well accounted for in the standard valence-only pseudopotential formalism.

DOI: 10.1103/PhysRevB.69.125212 PACS number(s): 71.15.−m, 71.15.Ap, 71.15.Dx, 71.10.−w

I. INTRODUCTION

Since the early applications of the GW method to real materials (see Refs. 1–3 and references therein), the pseudopotential plane-wave approach has been the method of choice due to its accuracy and technical simplicity. Recent advances in LAPW and linear muffin-tin orbital methodologies have allowed the implementation of ''all-electron'' applications of the GW method.[4–8] One common feature of such calculations, using standard level of approximation for the self-energy, is an underestimation of the electronic energy gap compared to experimental measurements, whereas pseudopotential-based calculations show very good agreement with experiment.[1,2,9] To explain this inconsistency, it was proposed that the pseudopotential approach does not correctly describe the effect of core orbitals in the self-energy corrections to the energy gaps, resulting in overestimated corrections.[6–8]

It is thus desirable to elucidate the effect of core orbitals in the quasiparticle band structure, and the preferred procedure is to perform a well converged all-electron calculation and compare its results with similarly converged pseudopotential-based calculations. Obviously, numerical precision should not be neglected. In this work, we explicitly include semicore orbitals in the pseudopotential plane-wave approach and calculate the quasiparticle energy gap for three semiconductors of technical importance: Si, Ge, and GaAs. The underlying description of the ground-state electronic structure is based on density-functional theory in the local density approximation (DFT/LDA).[10,11] Throughout this work, we are careful to converge all results systematically, and the final results are compared to previous pseudopotential results and to recent all-electron calculations. The paper is organized as follows. We outline the theoretical method in Sec. II. Results are presented in Sec. III and discussed in Sec. IV.

II. THEORETICAL METHOD

The pseudopotential formalism has two advantages that make it convenient for practical *ab initio* calculations. First, degrees of freedom due to core electrons are removed from the system, resulting in a description that contains only valence electrons. Second, valence electronic pseudowave functions are smooth in the vicinity of atomic sites as well as in the interstitial region. Such valence wave functions can be expanded to convergence easily in a relatively small basis set such as a plane-wave basis. On the other hand, one possible problem with this formalism is that using pseudopotentials, instead of the true electron-ion potential, may not fully describe all effects produced by interactions between the valence and core electrons. In order to address this issue in GW calculations, we include in the present study electrons from the outermost valence shell as well as those from the second outermost (''semicore'') atomic shell as active ''valence'' electrons in the pseudopotential formalism. Only interactions between these valence electrons and electrons from the deeper core shells are described by pseudopotentials. Taking Si as an example, its core now contains only $1s$ electrons. Electrons from the $2s$, $2p$, $3s$, $3p$, and $3d$ shells are all treated on equal footing in the subsequent calculations.

Apart from the atomic configuration, we follow the standard prescription for generating *ab initio* pseudopotentials.[12,13] Since the deeper core electrons have extremely large binding energy (e.g., 130 Ry for the $1s$ electrons in atomic Si), their interaction with valence electrons is expected to be much weaker than the already small interaction between the outermost valence and the semicore electrons.[14]

The Kohn-Sham DFT formalism within the LDA is used to solve for the ground-state electronic structure and to provide a starting point for the calculation of the electron self-energy. We follow closely the GW method as developed by Hybertsen and Louie.[1] In this method, the self-energy is given by the standard GW approximation,

$$\Sigma(\mathbf{r},\mathbf{r}';E)=i\int\frac{dE'}{2\pi}e^{-iE'0^+}G_0(\mathbf{r},\mathbf{r}';E-E')W_0(\mathbf{r},\mathbf{r}';E'), \tag{1}$$

TABLE I. Atomic parameters used to generate semicore pseudopotentials. Pseudowave functions were defined from the orbital with lowest principal quantum number at each angular momentum channel. Cutoff radii are given in units of Bohr radius.

	Reference configuration	r_{cut}			Local channel
		s	p	d	
Si	$2s^2 2p^6 3s^2 3p^{1.95} 3d^{0.05}$	0.40	0.35	0.40	d
Ge	$3s^2 3p^6 3d^{10} 4s^2 4p^{1.5} 4d^{0.1}$	0.50	0.50	0.50	s
Ga	$3s^2 3p^6 3d^{10} 4s^2 4p^1 4d^0$	0.50	0.50	0.50	s
As	$3s^2 3p^6 3d^{10} 4s^2 4p^2 4d^0$	0.50	0.50	0.45	p

where G_0 is the one-electron Green's function, calculated from LDA energy eigenvalues and eigenstates ($\delta = 0^+$ for occupied states, $\delta = 0^-$ for unoccupied states),

$$G_0(\mathbf{r},\mathbf{r}';E) = \sum_{n\mathbf{k}} \frac{\varphi_{n\mathbf{k}}(\mathbf{r})\varphi^{\star}_{n\mathbf{k}}(\mathbf{r}')}{E - \varepsilon^{DFT}_{n\mathbf{k}} - i\delta}. \quad (2)$$

The screened Coulomb interaction W_0 is calculated within the random-phase approximation (RPA) as $W_0 = [1 - VP_0]^{-1}V$, with V being the bare Coulomb interaction and the polarizability $P_0 = -iG_0G_0$.[15] This is the commonly employed level of approximation for GW calculations (i.e., neglecting self-consistency and vertex corrections),[2,3] and we specifically compare results from both pseudopotential and all-electron calculations at this particular level of approximation to help untangle the effects of core states.

The convolution integral in Eq. (1) is performed using the generalized plasmon-pole (GPP) model,[1] which enables one to distinguish two contributions to the self-energy: a screened exchange part (Σ_{sx}) arising from the poles of G_0 and a dynamical Coulomb interaction between an electron and the holelike charge distribution around it (Σ_{ch}) stemming from the poles of W_0. The former tends to increase the quasiparticle energy, after the bare exchange is excluded. The latter tends to decrease the quasiparticle energy. In particular, the dynamical contribution Σ_{ch} is highly sensitive to the number of bands n included in the calculation of the Green's function in Eq. (2). The final quasiparticle energy of a state $\varphi_{n\mathbf{k}}$ is given by (V_{xc} is the LDA exchange-correlation potential)

$$E^{qp}_{n\mathbf{k}} = \varepsilon^{DFT}_{n\mathbf{k}} + \langle \varphi_{n\mathbf{k}} | \Sigma(\mathbf{r},\mathbf{r}';E^{qp}_{n\mathbf{k}}) - V_{xc}(\mathbf{r}) | \varphi_{n\mathbf{k}} \rangle. \quad (3)$$

III. RESULTS

A semicore, nonrelativistic pseudopotential was generated for Si, using the Troullier-Martins scheme.[12] For Ga, Ge, and As, we constructed semirelativistic pseudopotentials using the Kerker scheme.[13] These choices resulted in stable, transferable pseudopotentials in the Kleinman-Bylander form, without ghost states.[16] A summary of atomic parameters is presented in Table I. A good expansion of electronic wave functions in a plane-wave basis was obtained using a cutoff energy of 700 Ry (600 Ry for Si), and the first Brillouin zone was sampled using a $4\times4\times4$ Monkhorst-Pack grid.[17] These numerical parameters ensure convergence in LDA energy eigenvalues to 0.01 eV or better. The Ceperley-Alder exchange-correlation potential is used.[11] For the lattice parameter, we used the experimental values: 5.43 Å, 5.65 Å, and 5.66 Å for Si, Ge, and GaAs, respectively. The polarizability was expanded in a plane-wave basis with an energy cutoff of 45 Ry (50 Ry for Si) and numerically inverted for the calculation of the screened Coulomb interaction W_0. Numerical precision in the calculation of the self-energy is 0.05 eV or better.

TABLE II. Band gaps of Si. All quantities in eV.

	E_{gap}	$\Gamma^v - \Gamma^c$	$\Gamma^v - X^c$
LDA present work			
	0.46	2.52	0.60
GW valence pseudopotential + CPP			
Shirley *et al.*[a]	1.13	3.28	1.31
GW present work			
	1.04	3.24	1.18
GW all-electron			
Hamada *et al.*[b]	1.01	3.30	1.14
Kotani and van Schilfgaarde[c]	0.89	3.12	
Ku and Eguiluz[d]	0.85	3.12	
Experiment[e]	1.17	3.35	1.3

[a]Reference 9.
[b]Reference 5.
[c]Reference 6.
[d]Non-self-consistent result from Ref. 7.
[e]Reference 23.

Table II shows some of the energy gaps obtained in the present approach for Si, compared with previous valence-only pseudopotential and all-electron calculations. Overall agreement between the present results (which explicitly include the effect of the semicore states) and experimental measurements is at the level of 0.1 eV, and discrepancies between our results and previous pseudopotential-based calculations of Ref. 9 are equally small. Recent all-electron calculations carried out at the same level of the GW approximation, however, systematically underestimate the minimum gap and the direct Γ-Γ gaps.[6,7] We find that the convergence of the self-energy with respect to the number of unoccupied bands included in G_0 in Eq. (2) is an important factor. In Fig. 1, we show the behavior of the calculated energy gap as function of the number of unoccupied bands, n_c, included in G_0. Convergence is typically very slow, and well-converged results require $n_c \geqslant 120$. Other energy transitions (Γ-X and Γ-Γ) show similar behavior and also approach the converged value from below. In contrast, the results of Ref. 7 were obtained with only $n_c = 24$ and are closer to our results at approximately the same value of n_c than to the converged results, as shown in Fig. 1. We see two possible explanations for this fact: lack of numerical convergence in the evaluation of the self-energy corrections in Ref. 7, or coincidence.

The energy gaps obtained for Ge are presented in Table III. Spin-orbit interactions are included as first-order perturbations.[18] A common feature of LDA-based calculations is the overlapping of the valence and conduction bands

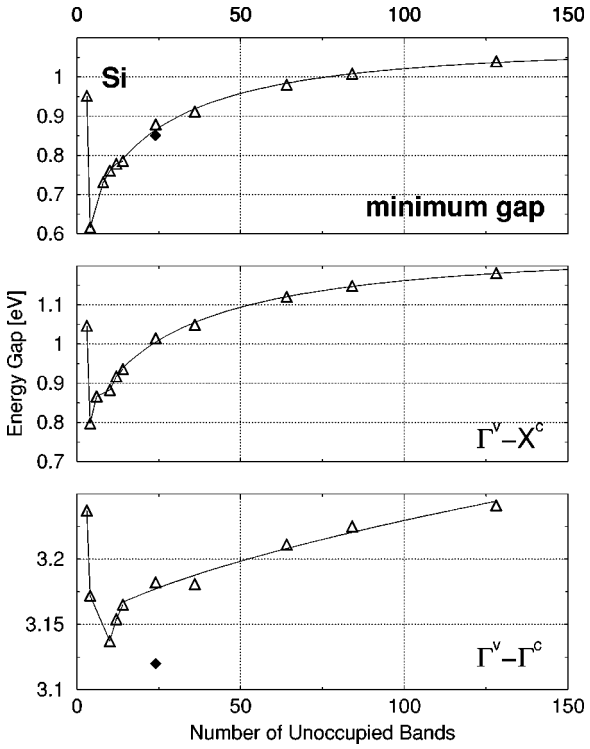

FIG. 1. Convergence of the minimum gap in Si as function of the number of unoccupied bands included in the calculation of the Green's function, Eq. (2) (open triangles). The solid line is a guide to the eye. Results obtained by Ku and Eguiluz (Ref. 7) are shown as black diamonds.

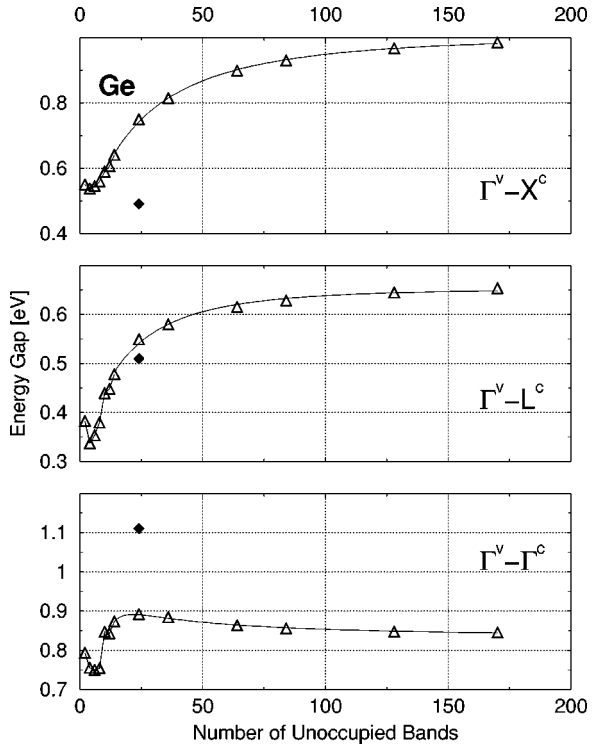

FIG. 2. Similar to Fig. 1 for Ge. Numerical values for the gaps Γ-Γ (lower panel), Γ-L (middle panel), and Γ-X (upper panel) are shown in open triangles. The full line is a guide to the eye. Results obtained by Ku and Eguiluz (Ref. 7) are shown in black diamonds.

at the Γ point. The inclusion of $3d$ electrons moves the conduction bands further down resulting in a sizable negative direct gap.[6,7,9,14,19] This feature is verified in our LDA calculation. As shown in Table III, self-energy corrections are responsible for an opening of the gap and the correct positioning of the minimum, indirect gap between points Γ and L in the Brillouin zone. Our GW results compare well with experiment, although the Γ-X gap still shows a large discrepancy. On the other hand, there are significant differences between the pseudopotential and all-electron gaps at this level of the GW approximation, in particular, regarding the Γ-X gap. As found in previous studies, the self-energy corrections are needed to give the correct band topology.

TABLE III. Band gaps of Ge. All quantities in eV.

	$\Gamma^v - L^c$	$\Gamma^v - \Gamma^c$	$\Gamma^v - X^c$
LDA present work			
	−0.04	−0.26	0.56
GW valence pseudopotential + CPP			
Shirley *et al.*[a]	0.73	0.85	1.09
GW present work			
	0.65	0.85	0.98
GW all-electron			
Kotani and van Schilfgaarde[b]	0.47	0.79	
Ku and Eguiluz[c]	0.51	1.11	0.49
Experiment[d]	0.74	0.90	1.3

[a]Reference 9.
[b]Reference 6 after inclusion of spin-orbit effects.
[c]Non-self-consistent results from Ref. 7.
[d]Reference 23.

Figure 2 shows direct and indirect energy gaps in Ge as function of the number of unoccupied bands included in the calculation of G_0. Whereas the indirect gaps Γ-L and Γ-X approach the converged value from below, the direct gap Γ-Γ approaches it from above. This particular convergence behavior arises from the fact that we are plotting differences of quasiparticle energies: taken individually, all quasiparticle energies $E^{qp}_{n,\mathbf{k}}$ converge monotonically from above, reflecting the attractive nature of the Coulomb-hole Σ_{ch} term.[1,2] Additionally, we note that in Ref. 7 the direct gap at Γ is strongly overestimated and the indirect gap Γ-X is underestimated by $\approx$0.8 eV. As expected, the same pattern of overestimation/underestimation is evident in Fig. 2 when we reduce the number of unoccupied bands from 170 to 24, which was the value used in Ref. 7.

GaAs shows behavior similar to Ge. Table IV summarizes our results. Agreement with experimental data is now within 0.15 eV, and all-electron calculations again underestimate the energy gaps. Regarding the convergence with n_c, Fig. 3 shows that the indirect gaps converge slowly from below. On the other hand, the direct Γ-Γ gap converges more quickly.

In all systems studied, we applied the GPP model[1] in two ways: (1) using only the valence charge density in the f-sum rule and (2) using a total charge density from both the va-

TABLE IV. Band gaps of GaAs. All quantities in eV.

	$\Gamma^v-\Gamma^c$	Γ^v-L^c	Γ^v-X^c
LDA present work			
	0.13	0.70	1.21
GW valence pseudopotential + CPP			
Shirley *et al.*[a]	1.42	1.75	1.95
GW present work			
	1.38	1.65	1.83
GW all-electron			
Kotani and van Schilfgaarde[b]	1.20	1.40	1.46
Experiment[c]	1.52	1.815	1.98

[a]Reference 9.
[b]Reference 6 after inclusion of spin-orbit effects.
[c]References 23,24.

lence and semicore bands. Method (1) is physically more realistic, and predicts an energy dependence of the inverse dielectric function that is consistent with the RPA. Method (2) implies that semicore electrons are able to screen electric fields as efficiently as valence electrons, which is not physical. Nevertheless, we find agreement between the two schemes to better than 0.1 eV in the converged energy gap for all systems studied. This is a consequence of cancellation of errors: in method (2), the plasma frequency is overestimated, and therefore the contributions Σ_{sx} and Σ_{ch} are overestimated in absolute value. Since they have opposite sign, the final quasiparticle energies are weakly affected. We also observe that method (2) shows slower convergence of energy gaps with respect to n_c, which is expected since the Σ_{ch} term is enhanced. All results presented in this article were obtained using the physically more appropriate method (1).

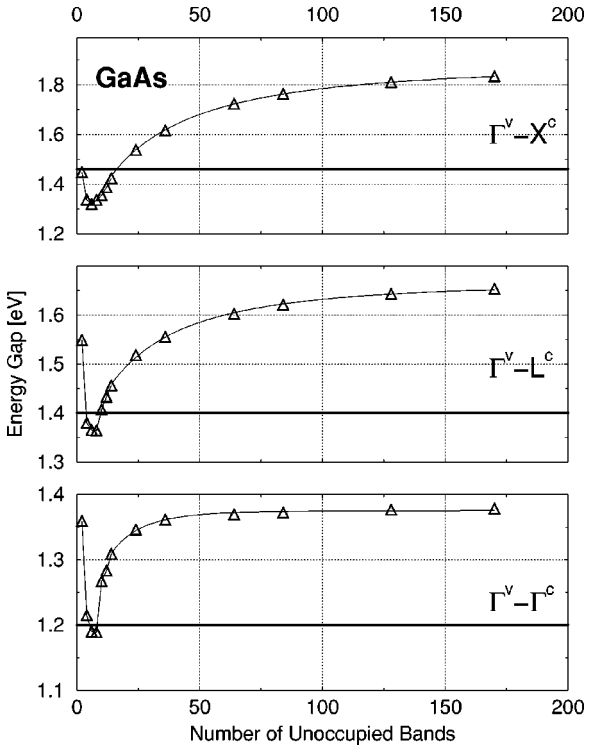

FIG. 3. Similar to Fig. 2 for GaAs. The thick horizontal line represents results obtained by Kotani and van Schilfgaarde (Ref. 6) (the number of bands used was not reported in this reference).

IV. DISCUSSION

As shown in Figs. 1–3, energy gaps in the GW approximation have a significant dependence on the number of unoccupied bands that is slow to converge in many cases. This fact was already reported in calculations using standard pseudopotential techniques.[2,20] We can analyze the physics of this convergence by examining the GW approximation in the static limit, the so-called COHSEX approximation.[1,2,15] Band gaps can be calculated more easily within the COHSEX approximation and show convergence behavior similar to the full dynamic calculation. Under this approximation, the Coulomb-hole term has a simple form in terms of the polarization potential, $W_{pol}(\mathbf{r},\mathbf{r}') \equiv W(\mathbf{r},\mathbf{r}',\omega=0)-V(\mathbf{r}-\mathbf{r}')$,

$$\Sigma_{COH}(\mathbf{r},\mathbf{r}') = \frac{1}{2}\delta(\mathbf{r}-\mathbf{r}')W_{pol}(\mathbf{r},\mathbf{r}') = \frac{1}{2}\sum_{n\mathbf{k}} \varphi_{n\mathbf{k}}(\mathbf{r})\varphi^{\star}_{n\mathbf{k}}(\mathbf{r}')W_{pol}(\mathbf{r},\mathbf{r}'), \quad (4)$$

where the second equality follows from completeness of the basis of eigenvectors. In actual calculations, this sum over bands n is always truncated, and the equality is violated. The Coulomb-hole energy evaluated at a given electronic state is calculated according to the expression

$$\langle m\mathbf{k}|\Sigma_{COH}|m\mathbf{k}\rangle = \frac{1}{2}\sum_{\mathbf{G}\mathbf{G}'\mathbf{q}} W_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) \times \sum_{n} [\mathcal{M}^{nm}_{\mathbf{G}}(\mathbf{k},\mathbf{q})]^{\star}\mathcal{M}^{nm}_{\mathbf{G}'}(\mathbf{k},\mathbf{q}), \quad (5)$$

where $W_{\mathbf{G}\mathbf{G}'}(\mathbf{q})$ are the coefficients in the plane-wave expansion of the polarization potential,[1,2] and we define

$$\mathcal{M}^{nm}_{\mathbf{G}}(\mathbf{k},\mathbf{q}) = \langle n\mathbf{k}-\mathbf{q}|e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}|m\mathbf{k}\rangle. \quad (6)$$

Physically, the summation in Eq. (4) describes virtual transitions produced when the quasiparticle induces a charge fluctuation around itself. Σ_{COH} is the energy associated to the interaction between the quasiparticle and the induced charge fluctuation. The matrix elements, Eq. (6), decay slowly as the energy difference between bands m and n increases. Monitoring the convergence of these matrix elements provides a good estimate of the relative error in the Σ_{COH}, but gauging the absolute convergence of the self-energy requires knowing W_{pol}, which depends on the physical system. From Eqs. (4)–(6), we see that matrix elements involving unoccupied bands are crucial in the calculation of the Σ operator. Therefore, the choice of basis set must be done carefully so that all occupied bands *and* the lowest unoccupied bands are accurately described. In this respect, a plane-wave basis set is expected to be more efficient than basis sets optimized for occupied bands.

Although we have not directly investigated the importance of imposing self-consistency in the calculation of the self-energy, this is an unsettled issue and we address it briefly below. Von Barth and Holm have investigated the effect of self-consistency in the electron-gas,[21] and concluded that restricted self-consistency has small but significant effect on the full bandwidth and in the satellite structure of the electron gas. On the other hand, full self-consistency gives a poor description of the satellite structure and the bandwidth is drastically increased. Inclusion of vertex corrections are expected to recover the good, non-self-consistent results, but calculation of vertex corrections is not a simple task even for the electron gas system.[2,21] Self-consistency has been recently applied to real materials,[7,22] and the valence bandwidth is also shown to increase when self-consistency is imposed. It appears that one must therefore include self-consistency and vertex corrections together in order to obtain a meaningful picture. Inclusion of vertex corrections and self-consistency is tangential for the purpose of our work, which is to compare pseudopotential-based and all-electron GW calculations and understand the role of semicore electrons.

V. CONCLUSION

We conclude that for the systems considered, the valence-only pseudopotential method does not suffer from large errors from the neglect of core states, as claimed in some all-electron calculations.[6,7] While semicore effects are negligible in Si, they are important if one aims at good quantitative agreement with experiment in Ge and GaAs.[2,3,14,19] However, discrepancies between pseudopotential and all-electron based GW calculations reported in recent works[6,7] may be explained by a lack of numerical convergence in the latter. Specifically, the self-energy calculated within the GW method has slow convergence with respect to the number of energy bands included in the calculation of the Green's function, as is demonstrated in this work and has been pointed out in the past.[2,20] This convergence behavior is present in the static limit to GW, the COHSEX approximation, and can be analyzed by comparing the COHSEX Coulomb-hole energy, Eq. (4), obtained with and without explicit summation over energy bands.

ACKNOWLEDGMENTS

We thank Andrew Canning and David Clatterbuck for fruitful discussions regarding the FLAPW method. This work was supported by National Science Foundation Grant No. DMR00-87088 and by the Director, Office of Science, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, U.S. Department of Energy under Contract No. DE-AC03-76SF00098. Computational resources have been provided by NSF at the National Partnership for Advanced Computational Infrastructure, and DOE at the National Energy Research Scientific Computing Center.

*Present address: Department of Applied Physics, Yale University, New Haven, CT 06520.

[1] M.S. Hybertsen and S.G. Louie, Phys. Rev. B **34**, 5390 (1986).
[2] W.G. Aulbur, L. Jönsson, and J.W. Wilkins, *Solid State Physics*, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 2000), Vol. 54, p. 1.
[3] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
[4] P.E. Blöchl, Phys. Rev. B **50**, 17 953 (1994).
[5] N. Hamada, M. Hwang, and A.J. Freeman, Phys. Rev. B **41**, 3620 (1990).
[6] T. Kotani and M. van Schilfgaarde, Solid State Commun. **121**, 461 (2002).
[7] W. Ku and A.G. Eguiluz, Phys. Rev. Lett. **89**, 126401 (2002).
[8] S. Lebègue, B. Arnaud, M. Alouani, and P.E. Blöchl, Phys. Rev. B **67**, 155208 (2003).
[9] E.L. Shirley, X. Zhu, and S.G. Louie, Phys. Rev. B **56**, 6648 (1997).
[10] R.M. Dreizler and E.K.U Gross, *Density Functional Theory: An Approach to the Quantum Many-body Problem* (Springer-Verlag, Berlin, 1990).
[11] J.P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[12] N. Troullier and J.L. Martins, Phys. Rev. B **43**, 1993 (1990).
[13] G.P. Kerker, J. Phys. C **13**, L189 (1980).
[14] M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **57**, 6485 (1998).
[15] L. Hedin and S. Lundqvist, *Solid State Physics* (Ref. 2), Vol. 23, p. 1.
[16] L. Kleinman and D.M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).
[17] H.J. Monkhorst and J.D. Pack, Phys. Rev. B **13**, 5188 (1976).
[18] M.S. Hybertsen and S.G. Louie, Phys. Rev. B **34**, 2920 (1986).
[19] F. Aryasetiawan and O. Gunnarsson, Phys. Rev. B **54**, 17 564 (1996).
[20] A. Fleszar and W. Hanke, Phys. Rev. B **56**, 10 228 (1997).
[21] U. von Barth and B. Holm, Phys. Rev. B **54**, 8411 (1996); **55**, 10 120 (1997); B. Holm and U. von Barth, *ibid.* **57**, 2108 (1998).
[22] W-D. Schöne and A.G. Eguiluz, Phys. Rev. Lett. **81**, 1662 (1998).
[23] Landolt-Börnstein, *Numerical Data and Functional Relationships in Science and Technology*, edited by K.-H. Hellwege, O. Madelung, M. Schulz, and H. Weiss, New Series Vol. III, Pt. 22a (Springer-Verlag, New York, 1987).
[24] P. Lautenschlager, M. Garriga, S. Logothetidis, and M. Cardona, Phys. Rev. B **35**, 9174 (1987).