

Use of the generalized gradient approximation in pseudopotential calculations of solids

Yu-Min Juan and Efthimios Kaxiras

Department of Physics and Division of Applied Sciences, Harvard University, Cambridge, Massachusetts 02138

Roy G. Gordon

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

(Received 19 October 1994)

We present a study of the equilibrium properties of *sp*-bonded solids within the pseudopotential approach, employing recently proposed generalized gradient approximation (GGA) exchange-correlation functionals. We analyze the effects of the gradient corrections on the behavior of the pseudopotentials and discuss possible approaches for constructing pseudopotentials self-consistently in the context of gradient-corrected functionals. The calculated equilibrium properties of solids using the GGA functionals are compared to the ones obtained through the local density approximation (LDA) and to experimental data. A significant improvement over the LDA results is achieved with the use of the GGA functionals for cohesive energies. For the lattice constant, the same accuracy as in LDA can be obtained when the nonlinear coupling between core and valence electrons introduced by the exchange-correlation functionals is properly taken into account. However, GGA functionals give bulk moduli that are too small compared to experiment.

There have been extensive efforts recently to improve the accuracy of density functional theory (DFT) (Ref. 1) by going beyond the local density approximation (LDA) (Ref. 2). Including gradient corrections to LDA represents a promising scheme that is conceptually simple.³⁻¹⁰ This approach is referred to as the generalized gradient approximation, or GGA. Calculations using different gradient-corrected functionals have been performed to test the applicability of this approach on a variety of systems.¹¹⁻¹⁷

In an earlier work,¹⁷ we investigated one GGA functional recently proposed by Perdew and Wang (PW91) (Refs. 6 and 9) in the calculations of both atoms and solids. We showed that by simply combining the PW91 functional with the pseudopotential approach leads to lattice constants for solids, such as simple metals and semiconductors, that are larger than experiment, and the percentage errors are significantly larger than those obtained from LDA. Results from all-electron calculations^{13,16} for systems other than the ones we considered, such as transition metals, suggest that there is no such significant increase of the error in lattice constants when the PW91 functional is used. In order to understand this difference between the results obtained from the pseudopotential and the all-electron calculations, we first examined the charge distribution of individual atoms. We considered both PW91 and the more recently proposed functional by Lacks and Gordon (LG).¹⁰ The reasons for choosing these two functionals are the following: (1) They are among the most recent additions to the list of proposed gradient corrections and are intended to give better results compared to earlier attempts. (2) The two functionals have similar expressions which simplifies computational implementation. (3) The Lacks-Gordon exchange functional is produced by fitting to exact results, while the PW91 functional is derived from first principles; comparison between these two func-

tionals may provide insight for a more accurate approach.

In studying the charge distributions of individual atoms, we have found it instructive to evaluate the following quantity:

$$C_{nl}(R) = 4\pi \int_0^R r^2 dr [|\phi_{nl}^{\text{LDA}}(r)|^2 - |\phi_{nl}^{\text{GGA}}(r)|^2]. \quad (1)$$

$C_{nl}(R)$ is the difference between the charge enclosed within a sphere of radius R around the nucleus calculated from the LDA and the GGA functional for each single electron orbital nl . In Fig. 1 we display $C_{nl}(R)$ for the $nl = 2p, 3s$, and $3p$ orbitals of Si, and the $nl = 2p, 3s$, $3p$ orbitals of Na. The more positive $C_{nl}(R)$ is, the more charge has been pushed outside the region contained by the sphere of radius R in the GGA calculation compared to the charge obtained by LDA.

In order to see the difference in the physically important range, we used the experimentally measured bond length as the unit along the x axis in Fig. 1. As is obvious from this figure, there is almost no difference in the charge distribution between the LDA and GGA results for the $2p$ core orbital of Si. For the $3s$ and $3p$ valence orbitals of Si, in the neighborhood of the bonding region, substantial charge has been pushed away from the nucleus in both the PW91 and the LG calculations, relative to the LDA results. A similar situation occurs in the case of Na as is shown in Fig. 1.

These comparisons indicate a weaker interaction between the valence electrons of the atom and the ion when using GGA functionals as opposed to LDA, due to the spreading out of the valence charge in GGA calculations. In an approach that simply replaces the effects due to the ion and the core electrons with a pseudocore that is constructed to reproduce the results of the all-electron calculation, this will unavoidably lead to a weaker interaction between the valence electrons and the pseudocore.

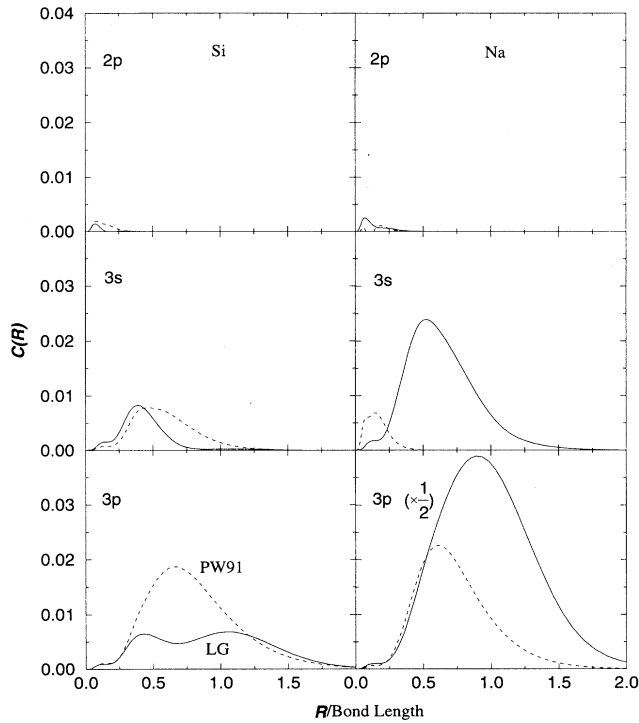


FIG. 1. Integrated charge difference $C_{nl}(R)$ [see Eq. (1)] calculated with PW91 (dashed lines), and LG (solid lines) for $nl = 2p, 3s, 3p$ orbitals of Si, and $nl = 2p, 3s, 3p$ orbitals of Na. The results for the $3p$ orbital of Na have been divided by a factor of 2 so that they can be displayed on the same scale as the results for Si.

Since in the pseudopotential framework the properties of solids are determined by the interaction between valence electrons and the pseudocore, the tendency for the valence charge to be pushed away from the nucleus when GGA functionals are used leads to a softer solid, characterized by a larger equilibrium lattice constant and a smaller bulk modulus than LDA.

This observation points to the necessity of properly taking into account the nonlinear coupling between the valence and core electrons in the exchange-correlation functional within the pseudopotential approach, when GGA functionals are used. To examine the effects due to the inclusion of gradient corrections, we first consider the behavior of the quantity s which is used in the definition of GGA functionals in addition to the charge density n . s is defined as a function of the charge density n and its gradient ∇n

$$s = |\nabla n| / 2(3\pi^2)^{1/3} n^{4/3}. \quad (2)$$

Although the charge density n can be separated into the core charge density n_c and the valence charge density n_v so that $n = n_c + n_v$, such a separation cannot be written for s due to the nonlinearity of the expression of Eq. (2). We use Si as an example to further illustrate this point. We display in Fig. 2 both the core and valence charge density for the Si atom, calculated with the LDA and

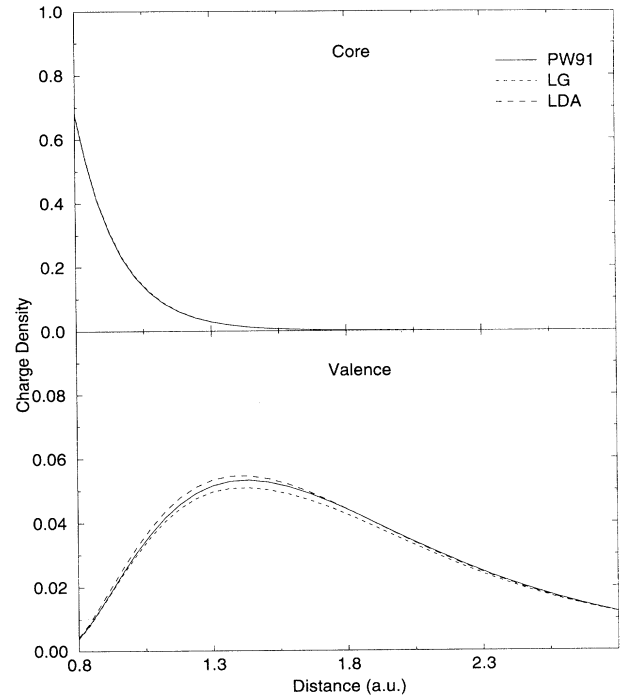


FIG. 2. The core (upper panel) and valence (lower panel) charge density of the silicon atom calculated with LDA and GGA functionals as a function of distance from the nucleus. Note the different density scales in the two panels.

the GGA functionals, respectively. There is evidently no significant difference between the results of these two calculations as far as the overlap between charge density is concerned. Since the results from the PW91 and LG functionals tend to be very similar (see, for example, the comparison in Fig. 1), from now on we will only present results obtained with PW91. Figure 3(a) displays the quantity

$$\Delta V_{xc} = V_{xc}[n_c] + V_{xc}[n_v] - V_{xc}[(n_c + n_v)] \quad (3)$$

which is a measure of the nonlinearity of the exchange-correlation potential. There is a substantial increase in the nonlinearity of the exchange-correlation potential in the region where the overlap between the core and valence charge is not negligible (compare with Fig. 2). This increase in nonlinearity is due to the fact that the variables used in the expression for the GGA functionals are not separable in terms of the valence and core parts. For example, we show in Fig. 3(b) the values of s_c , s_v , and s which correspond to the value of the quantity defined in Eq. (2) calculated with the core, valence, and total charge density, respectively. Therefore, the inclusion of gradient corrections seems to increase the effects of coupling between the valence and core electrons.

As was discussed in the previous paragraphs, the simple unscreening of the pseudopotential by

$$V_{\text{ion}}^{\text{PS}} = V_{\text{screened}}^{\text{PS}} - V_H[n_v] - V_{xc}[n_v], \quad (4)$$

does not give satisfactory results for the properties of

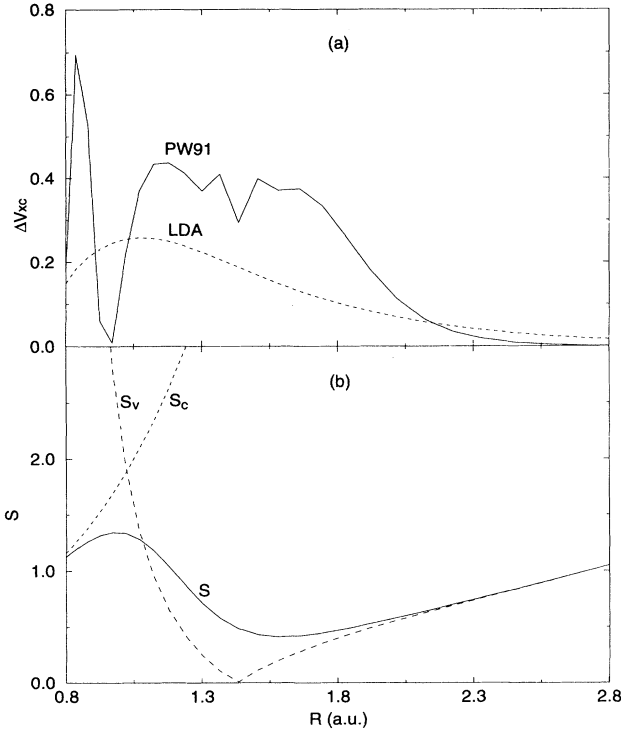


FIG. 3. (a) The quantity ΔV_{xc} [see Eq. (3)] for the silicon atom calculated with LDA and PW91 as a function of distance from the nucleus. (b) The values of s [see Eq. (2)] calculated from the core (s_c), the valence (s_v), and the total charge density (s) for the silicon atom as a function of distance from the nucleus.

solids; here, $V_{\text{screened}}^{\text{ps}}$ is the screened pseudopotential obtained by directly inverting the Schrödinger equation (see, e.g., Ref. 17), V_H and V_{xc} are the Hartree and the exchange-correlation potentials, respectively. This approximation implicitly assumes the linearization of the exchange-correlation functional. As was pointed out by Louie *et al.*,¹⁸ a more consistent approach is to include the core charge density in the unscreening. That is, instead of taking out only the $V_{xc}[n_v]$ part in the unscreening procedure as in Eq. (4), $V_{\text{ion}}^{\text{ps}}$ should be defined as

$$V_{\text{ion}}^{\text{ps}} = V_{\text{screened}}^{\text{ps}} - V_H[n_v] - V_{xc}[(n_c + n_v)], \quad (5)$$

where n_c is a rigid core charge density, constructed from a reference atomic system. Since the core electrons are not included in the pseudopotential calculation, whenever the exchange-correlation energy and potential are needed, the full charge density $n = n_c + n_v$ must be used. This procedure is exact within the rigid core approximation, but it would require a very large number of plane waves to describe the core charge density accurately, and one loses the advantages of using the pseudopotential formalism by adopting this approach. So even though it is theoretically correct, it is not practical from a computational point of view. Therefore, it is necessary to make some approximation in order to obtain a practical computational scheme.

In the present paper, we follow the partial core prescription proposed in Ref. 18. The full core charge den-

sity is replaced with an artificial core charge density \tilde{n}_c defined as

$$\tilde{n}_c = \begin{cases} \frac{A \sin Br}{r}, & r \leq r_c \\ n_c, & r > r_c. \end{cases} \quad (6)$$

The parameters A and B are determined by the requirement that the value of \tilde{n}_c and its derivative with respect to the radius r be exactly the same as those of the real core charge density n_c at the cutoff radius r_c . We have found that in order to capture the nonlinear coupling between the core and the valence electrons for the case of the PW91 functional, it is necessary to use r_c smaller than what was suggested in Ref. 18 for LDA calculations. In our calculations, r_c is chosen as the radius where the core charge density n_c is 6–7 times larger than the valence charge density n_v . In Ref. 18 (which dealt with LDA calculations), r_c was chosen as the radius where the core charge density n_c was 2–3 times larger than the valence charge density. It is worth mentioning that the pathological oscillatory behavior of the PW91 exchange-correlation potential near the nuclei, which causes problems in creating smooth pseudopotentials during the unscreening procedure,^{13,17} is automatically eliminated by using the partial core correction.

As an illustration of how this approach works, we consider four *sp*-bonded solids in their ground state phase: Si (diamond), Ge (diamond), GaAs (zinc blende), and Al (fcc). We construct self-consistent pseudopotentials as described in Ref. 17 with the partial core prescription for the unscreening procedure discussed above. For the LDA calculations we use the exchange-correlation potential of Ceperley and Alder as parametrized by Perdew and Zunger¹⁹ and norm-conserving pseudopotentials from Bachelet, Hamann, and Schlüter.²⁰ We use a plane-wave basis for the expansion of the wave function of valence electrons: the highest kinetic energy of the plane waves in the basis is 16 Ry. For reciprocal space integrations, 29 special k points in the irreducible Brillouin zone are used for the diamond structure and the zinc-blende structure, and 213 special k points for the fcc structure. The gradient and the Laplacian of the density, which are needed for the GGA functionals considered here, were obtained through fast Fourier transforms, with minimal increase in CPU time (less than 3%). The calculated energy vs volume results are fitted to the Birch-Murnaghan equation of state.²¹ The equilibrium properties are then derived from the equation of state curves. The cohesive energy is taken to be the total energy difference between the solid in equilibrium and the isolated atom. Spin polarization effects on the free atom energy are taken into account with the empirical formula $\Delta E_p = -0.18n_p^2$,^{22,23} where $n_p = n_\uparrow - n_\downarrow$ with n_\uparrow the number of electrons having spin up and n_\downarrow the number of electrons with spin down. Spin polarization is expected to have negligible effects on the total energy E_0 of nonmagnetic solids.

The calculated ground state properties using LDA and PW91 are summarized and compared to experimental data in Table I. The results from PW91 represent a substantial improvement of the overbinding problem of LDA: the cohesive energies are in better agreement with experiment for all the solids we have considered. This improve-

TABLE I. Lattice constant (a_0 in Å), bulk modulus (B in GPa), cohesive energy (E_{coh} in eV/atom), calculated from LDA and PW91: (1) the LDA results, (2) the PW91 results obtained from a pseudopotential calculation without partial core correction, (Ref. 17), (3) the present PW91 results by using a pseudopotential with partial core correction, (4) the PW91 results from an all-electron linearized augmented plane-wave (LAPW) calculation (Ref. 24), and (5) the experimental values. The values in parentheses are percent differences from the experimental values.

	Si			Al		
	a_0	B	E_{coh}	a_0	B	E_{coh}
(1)	5.38 (-0.9%)	96.57 (-2%)	5.38 (+15.9%)	3.93 (-2.9%)	87.65 (+21%)	4.14 (+22.1%)
(2)	5.59 (+2.9%)	85.22 (-14%)	4.64 (+0.2%)	4.25 (+4.9%)	61.07 (-15%)	3.22 (-5.0%)
(3)	5.45 (+0.4%)	83.51 (-15%)	4.40 (-5.1%)	4.03 (-0.5%)	79.25 (+10%)	3.45 (+1.7%)
(4)	5.50 (+1.3%)	83.00 (-16%)				
(5)	5.43	98.80	4.63	4.05	72.20	3.39
	Ge			GaAs		
	a_0	B	E_{coh}	a_0	B	E_{coh}
(1)	5.57 (-1.6%)	74.72 (-3%)	4.53 (+17.7%)	5.51 (-2.5%)	77.09 (+2%)	8.58 (+31.6%)
(2)	5.92 (+4.7%)	58.12 (-25%)	3.71 (-3.6%)	5.92 (+4.8%)	54.59 (-28%)	6.43 (-1.4%)
(3)	5.73 (+1.2%)	58.04 (-24%)	3.41 (-11.4%)	5.76 (+1.9%)	66.28 (-12%)	6.00 (-8.6%)
(4)	5.75 (+1.6%)	61.00 (-21%)		5.74 (+1.5%)	65.00 (-14%)	
(5)	5.66	77.20	3.85	5.65	75.57	6.52

ment in cohesive energies can be attributed to a large extent to the fact that PW91 gives a more accurate atomic energy than LDA.¹⁷ For the equilibrium lattice constant, the value obtained from PW91 is consistently larger than the LDA results. In the case of Al, this makes the value obtained from PW91 closer to experiment than the LDA result. For Si, Ge, and GaAs, the results obtained from LDA and PW91 are of the same accuracy compared to experiment. PW91 tends to overcorrect the LDA results and gives an overestimate for the equilibrium lattice constant of these systems. For the bulk modulus, the values obtained from PW91 are smaller than the LDA results. While this leads to a better result for Al, the bulk moduli we obtained for the three semiconductors are significantly underestimated (by -12% to -25% compared to experiment). Similar observations for earlier gradient-corrected functionals have been reported.^{11,12} Finally, we compare our results to recent all-electron, linearized augmented plane-wave (LAPW) (Ref. 24) total energy and electronic structure calculations, with the same GGA functional as in our work. It is obvious from the comparison of Table I that the present pseudopotential calculation results with the partial core correction represent a significant improvement over the results without the partial core correction, and agree very well with the all-electron LAPW calculation results. The remaining discrepancy is probably due to the relaxation of the core electrons which is not allowed in the pseudopotential calculation.

For the electronic structure, we compare in Table II the band gap predicted by LDA and PW91 at both the experimentally measured lattice constant and the theoretical equilibrium lattice constant. An overall improvement that brings the values closer to the experimental data at the experimental lattice constant is found for all three semiconductors we have considered, although the magnitude of the improvement depends on the material. There is no consistent improvement for the band gaps at the theoretical equilibrium lattice constant. We note here that this is simply a comparison between LDA and PW91 as different approximations to the exchange-correlation

functionals. The well-known inability of density functional theory to reproduce accurately band gaps in semiconductors and insulators is much more complicated and related to the intrinsic discontinuity of the exchange-correlation functional,²⁵ which is not represented by either of the two approximations used here. Good agreement for the band gap values nevertheless can be obtained by using DFT/LDA wave functions and solving the self-energy operator equations, within the so-called GW approximation.²⁶

In conclusion, we showed that it is essential to take into account the core-valence coupling in the pseudopotential calculations when using GGA exchange-correlation functionals. To this end, we have found that the partial core prescription of Louie *et al.*¹⁸ is most appropriate when using a plane-wave basis. We considered the structural properties of Si, Ge, GaAs, and Al using both LDA and PW91. We found that PW91 gives consistently better cohesive energies than LDA. We also demonstrated that for the lattice constant the same accuracy as in LDA can be obtained with GGA, as long as the nonlinearity of the gradient-corrected functional is properly taken into account. For the semiconductors we considered, the bulk moduli obtained with the use of GGA functionals represent significant underestimates of the experimental results. The PW91 functional does give a better description for the equilibrium properties of Al. We conclude that a further search may be needed for an exchange-

TABLE II. Band gap in eV calculated from LDA and PW91; a_0^{th} is the equilibrium lattice constant predicted by theory and a_0^{exp} is the experimentally measured lattice constant.

	Si		Ge		GaAs	
	at a_0^{th}	at a_0^{exp}	at a_0^{th}	at a_0^{exp}	at a_0^{th}	at a_0^{exp}
LDA	0.49	0.54	0.30	0.02	1.03	0.58
PW91	0.65	0.65	0.43	0.69	0.73	1.18
Expt.		1.17		0.79		1.52

correlation functional that is consistently better for all solids.

In view of the above results, one may inquire what are the physical situations in which the use of GGA functionals can provide significant improvements over LDA results. Recently, calculations have been reported for H₂ dissociation on a Cu(111) surface with the LDA and the GGA.^{27,28} The GGA results for this system represent significant improvements over the LDA results. It has also been demonstrated that the GGA gives results in better agreement with experiments than the LDA for finite systems (atoms and molecules) and metallic surfaces.¹³⁻¹⁵ It is, therefore, expected that the GGA will give, in general,

a better description for the interaction between molecules and other molecules or solid surfaces. The reason that the GGA should give better results for these interactions can be attributed to the fact that a substantial part of the interactions in the these systems are related to the tails of the electronic wave functions, where the GGA gives a more accurate description than the LDA.

This work was supported by the Materials Research Laboratory of Harvard University which is funded by National Science Foundation Grant No. DMR 89-20490. The calculations were carried out at the Cornell National Supercomputer Facility.

-
- ¹ P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
 - ² W. Kohn and L. Sham, *Phys. Rev.* **140**, A1133 (1965).
 - ³ D.C. Langreth and M.J. Mehl, *Phys. Rev. B* **28**, 1809 (1983); D.C. Langreth, in *Many-Body Phenomena at Surfaces*, edited by D.C. Langreth and H. Suld (Plenum, New York, 1984).
 - ⁴ *Theory of the Inhomogeneous Electron Gas*, edited by S. Lundqvist and N.H. March (Plenum, New York, 1983).
 - ⁵ C.D. Hu and D.C. Langreth, *Phys. Scr.* **32**, 391 (1985).
 - ⁶ J.P. Perdew, *Phys. Rev. Lett.* **55**, 1665 (1985); J.P. Perdew and Y. Wang, *Phys. Rev. B* **33**, 8800 (1986).
 - ⁷ A.D. Becke, *Phys. Rev. A* **38**, 3098 (1988); F.W. Kutzler and G.S. Painter, *Phys. Rev. Lett.* **59**, 1285 (1987).
 - ⁸ M. Norman and D.D. Koelling, *Phys. Rev. B* **28**, 4357 (1983).
 - ⁹ J.P. Perdew, in *Electronic Structure of Solids '91*, edited by P. Ziesche and H. Eschrig (Akademie Verlag, Berlin, 1991).
 - ¹⁰ D.J. Lacks, and R.G. Gordon, *Phys. Rev. A* **47**, 4681 (1993).
 - ¹¹ B. Barbiellini, E.G. Moroni, and T. Jarlborg, *J. Phys. Condens. Matter* **2**, 7597 (1990).
 - ¹² A. García, C. Elässer, J. Zhu, S. Louie, and M.L. Cohen, *Phys. Rev. B* **46**, 9829 (1992).
 - ¹³ G. Ortiz and P. Ballone, *Phys. Rev. B* **43**, 6376 (1991).
 - ¹⁴ J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, and C. Fiolhais, *Phys. Rev. B* **46**, 6671 (1992).
 - ¹⁵ B.G. Johnson, P.M.W. Gill, and J.A. Pople, *J. Chem. Phys.* **98**, 5612 (1993).
 - ¹⁶ V. Ozolins and M. Korling, *Phys. Rev. B* **48**, 18304 (1993).
 - ¹⁷ Y.M. Juan and E. Kaxiras, *Phys. Rev. B* **48**, 14944 (1993).
 - ¹⁸ S.G. Louie, S. Froyen, and M.L. Cohen, *Phys. Rev. B* **26**, 1738 (1982).
 - ¹⁹ D.M. Ceperley and B.J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980); J. Perdew and A. Zunger, *Phys. Rev. B* **23**, 5048 (1981).
 - ²⁰ G.B. Bachelet, D.R. Hamann, and M. Schlüter, *Phys. Rev. B* **26**, 4199 (1982).
 - ²¹ F.D. Murnaghan, *Proc. Natl. Acad. Sci. U.S.A.* **30**, 244 (1944).
 - ²² O. Gunnarson, B.I. Lundqvist, and J.W. Wilkins, *Phys. Rev. B* **10**, 1319 (1974).
 - ²³ J. Ihm and J.D. Joannopoulos, *Phys. Rev. B* **24**, 4191 (1981).
 - ²⁴ C. Filippi, D.J. Singh, and C.J. Umrigar, *Phys. Rev. B* **50**, 14947 (1994).
 - ²⁵ J. Perdew and M. Levy, *Phys. Rev. Lett.* **51**, 1884 (1983); L. Sham and M. Schlüter, *ibid.* **51**, 1888 (1983); M. Lannoo, M. Schlüter, and L.J. Sham, *Phys. Rev. B* **32**, 3890 (1985).
 - ²⁶ L. Hedin, *Phys. Rev.* **139**, A796 (1965); M. Hybertsen and S. Louie, *Phys. Rev. Lett.* **55**, 1418 (1985); *Phys. Rev. B* **34**, 5390 (1986); R. Godby, M. Schlüter, and L. Sham, *Phys. Rev. Lett.* **56**, 2415 (1986); *Phys. Rev. B* **37**, 10159 (1988).
 - ²⁷ B. Hammer, M. Scheffler, K.W. Jacobsen, and J.K. Nørskov, *Phys. Rev. Lett.* **73**, 1400 (1994).
 - ²⁸ J.A. White, D.M. Bird, M.C. Payne, and I. Stich, *Phys. Rev. Lett.* **73**, 1404 (1994).