Efficient Band Gap Prediction for Solids

M. K. Y. Chan[1] and G. Ceder[2]

[1]Physics/Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[2]Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

(Received 27 March 2010; published 5 November 2010)

An efficient method for the prediction of fundamental band gaps in solids using density functional theory (DFT) is proposed. Generalizing the Delta self-consistent-field (ΔSCF) method to infinite solids, the Δ-sol method is based on total-energy differences and derived from dielectric screening properties of electrons. Using local and semilocal exchange-correlation functionals (local density and generalized gradient approximations), we demonstrate a 70% reduction of mean absolute errors compared to Kohn-Sham gaps on over 100 compounds with experimental gaps of 0.5–4 eV, at computational costs similar to typical DFT calculations.

DOI: 10.1103/PhysRevLett.105.196403 PACS numbers: 71.15.Dx, 71.15.Mb, 71.20.−b

Density functional theory (DFT) [1], in the Kohn-Sham implementation [2] with local density (LDA [2]) or generalized gradient (GGA [3]) approximations for the exchange-correlation (XC) functional, has been successfully applied to deduce structural, electronic, magnetic and other properties of a myriad of condensed matter systems. However, the well-known "band gap problem", in which DFT in LDA/GGA fails to correctly predict the energy gaps between occupied and unoccupied states, is a hindrance to research in fields including semiconductors, optical and photovoltaic materials, and thermoelectrics.

In a typical DFT calculation, the Kohn-Sham gap E_{KS}, i.e., difference between eigenvalues of lowest unoccupied and highest occupied eigenstates, is identified as the band gap. Typically, E_{KS} underestimates band gaps (E_{Exp}) of solids by 30%–100% [4]. To illustrate the problem, we plot in Fig. 1 E_{KS} vs E_{Exp} (red dots) for a test set of over 100 semiconductor compounds [5] with room-temperature E_{Exp} of 0.5–4 eV, a range of particular interest for applications. Not only are the errors large, there is no discernible correlation between E_{KS} and E_{Exp}.

The underestimation of E_{Exp} by E_{KS} in local and semilocal functionals has been attributed to their inherent lack of derivative discontinuity [6] and delocalization error [7]. It is often claimed that the band gap is an excited-state property and therefore cannot be described by ground state DFT [8]. Much effort has been devoted to solving the DFT band gap problem both within and in addition to the Kohn-Sham formalism, e.g., through the *GW* approximation [9], time-dependent DFT [10], exact exchange [11], hybrid and screened-hybrid [12] functionals, and modified Becke-Johnson (MBJ) potentials [13].

In this Letter, we report an efficient method for predicting band gaps for solids using computationally efficient local and semilocal functionals. The method is drawn upon the well-known Delta self-consistent-field (ΔSCF) [14] method, and will be referred to as Δ-sol. The predicted gaps with Δ-sol are plotted against the experimental values (blue crosses) in Fig. 1. The underlying physical principle of Δ-sol, as in *GW* and screened-hybrid functionals, is a consideration of the dielectric screening properties of electrons. In the implementation of Δ-sol, as in hybrid functionals and MBJ, one or two parameters are fixed to minimize errors in predicting E_{Exp} for a test set of compounds. Like MBJ and unlike *GW* and hybrid functionals, Δ-sol is as computationally expedient as standard DFT. Unlike MBJ, any existing DFT code and functional can be used in conjunction with Δ-sol without recoding.

The fundamental gap E_{FG} of a system is defined as the energy required to create an unbound electron-hole pair:

$$E_{\mathrm{FG}} = E(N+1) + E(N-1) - 2E(N), \qquad (1)$$

where N is the number of electrons in the system. The evaluation of E_{FG} from (1) by explicit calculations of energies of the system with N, $N+1$ and $N-1$ electrons, i.e., the ΔSCF method, produces reasonable results for atoms and molecules [7,14]. The problem that arises, in

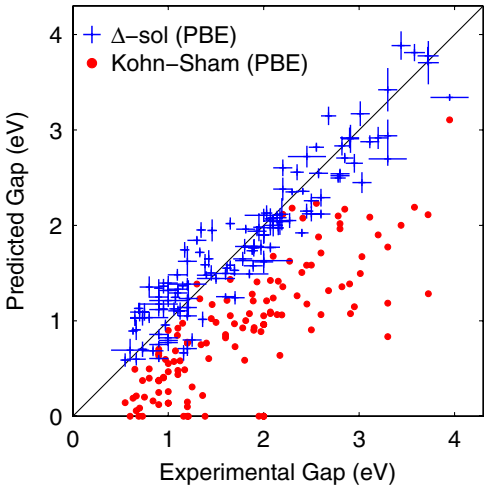

FIG. 1 (color online). Kohn-Sham (dots) and Δ-sol (crosses) vs experimental [5] gaps. Cross sizes represent uncertainties.

applying (1) to a solid, is in determining N. If one takes $N \to \infty$ for an infinite solid, E_{FG} reverts to E_{KS} [6,14].

In order to apply (1) to a solid, it is instructive to consider the addition and removal of charges to an electron gas. Figure 2 shows the integrated screening charge N_s for the screening of a static unit charge by a homogeneous electron gas (HEG), found using the RPA static dielectric function by Langer and Vosko [15]. Beyond a distance of $5–7k_f^{-1}$, where $k_f = (3\pi^2 N/V)^{1/3}$ is the Fermi wave vector, N_s saturates, indicating complete screening. Figure 2 also shows integrated screening charges, from *ab initio* **q**-dependent RPA dielectric functions computed [16] for cubic Si, GaAs, C, and ZnS, which saturate at a similar range. The charge distribution of electrons due to mutual screening is described by the pair distribution function $g(\mathbf{r})$ or XC hole, which is related to the dielectric function in a way similar to N_s [17]. The spherically averaged $g(r)$ in a HEG from quantum Monte Carlo [18] or LDA [19] has a form similar to Fig. 2 and also saturates at $5–7k_f^{-1}$.

LDA owes its success in part to the fact that it satisfies the XC-hole sum rule [20,21]. In addition, the spherical average and the extent of the XC hole in LDA are close to the exact results [22]. These are integral quantities that are not sensitive to the precise details of $g(\mathbf{r})$. To obtain the fundamental gap, we are similarly interested in an integral quantity, namely, the total energy due to the added or removed electron and its screening charge distribution. Based on the attribution of band gap errors in LDA/GGA to delocalization error [7], we propose that as long as we avoid the delocalization error by confining the added charge to a volume that is commensurate with the range of the screening effects, the integrated energies thus obtained will be reasonably correct. In other words, we propose that the appropriate N for which to evaluate E_{FG} from (1) in an infinite solid is the number of electrons contained within the extent of the XC hole, R_{XC}. Since $R_{\mathrm{XC}} \simeq 5–7k_f^{-1}$, the corresponding number of electrons inside the XC hole is estimated to be $N^* = \beta \frac{R_{\mathrm{XC}}^3}{3\pi^2} \sim 20–90$, where $\beta \simeq 4–8$ is a geometric factor depending on the shape of the units used to tile all space. We term the evaluation of band gaps from (1) by adding and removing one electron per screening volume, i.e., one per N^* electrons, the Δ-sol method. We propose that Δ-sol is valid, although possibly with a different value of N^*, for any XC functional which gives reasonably accurate total energies.

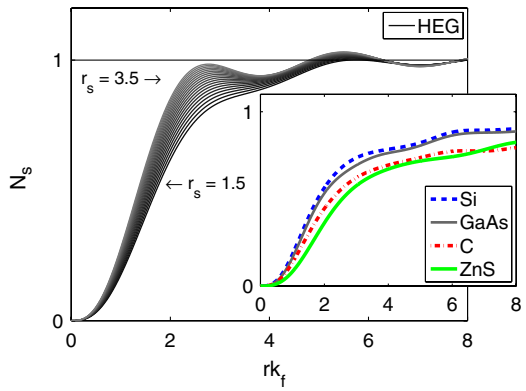

FIG. 2 (color online). Integrated screening charge N_s for a unit charge in a homogeneous electron gas (HEG) and selected semiconductors (inset), vs the dimensionless distance rk_f. Different HEG curves correspond to different values of the density parameter r_s, in a range (1.5–3.5) typical for solids.

We test the proposal that (1) can be used with $N = N^* \in [20, 90]$ to obtain the experimental band gap on the aforementioned test set. We perform DFT total-energy calculations at various electron numbers with different local and semilocal XC functionals (LDA, PBE [23] and AM05 [24]) using the plane-wave DFT code VASP and PAW potentials [25]. Because of the dependence of band gaps on lattice parameters, we use room-temperature experimental lattice constants. We use dense ($\sim 10^4$ Å$^3/V$, V = cell volume) Γ-centered k-point grids, and tetrahedron smearing method with Blöchl corrections [26]. To obtain accurate energies by reducing the self-interaction error [27], the DFT + U [28] scheme with $U = 3$ eV is used for oxides and halides containing partially filled d shells.

For each compound, we determine the number of valence electrons in the unit cell, N_0. For main group elements, the number of electrons that each ion contributes to N_0 is determined according to the usual octet rule. For transition metals, all outermost s and d electrons are counted. The number of valence electrons assigned this way is independent of the pseudopotentials used.

To add or remove one per N valence electrons, the number of electrons to add/remove from a unit cell with N_0 valence electrons is $n = N_0/N$. In most cases n is not an integer. We calculate the energies $E(N_0)$, $E(N_0 + n)$, and $E(N_0 - n)$, from which we obtain

$$E_{\mathrm{FG}} = [E(N_0 + n) + E(N_0 - n) - 2E(N_0)]/n. \qquad (2)$$

For $N = N^*$, this procedure is equivalent to adding and removing one electron per screening volume.

As N is increased, the calculated E_{FG} decreases monotonically, as illustrated in Fig. 3. The best value of N, i.e., N^*, is determined for each functional such that the mean absolute error between E_{FG} and E_{Exp} across the test set is minimized. The uncertainty in N^* is determined by (i) leave 20% out cross validation, (ii) equalization of the mean uncertainties in E_{FG} and E_{Exp}, and (iii) allowing the mean absolute error ϵ to vary by $\sigma/\sqrt{M}$, where σ is the standard deviation in the absolute error and M is the number of compounds. The three methods give similar uncertainties in N^*. The values of N^* and uncertainties by method (iii) are shown in Table I. The key results are that, for all three functionals tested, $N^* \in [20, 90]$ as proposed, and the Δ-sol method reduces Kohn-Sham gaps errors by an average of 70%.

We note from Table I that N^*_{best} only varies slightly with the XC functional and subgroup of compounds. The constancy of N^* across different functionals and the fact that

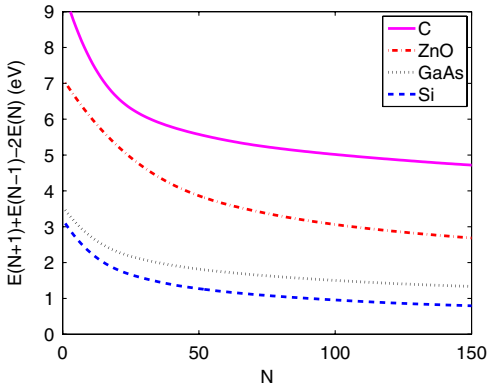

FIG. 3 (color online). The calculated E_{FG} (PBE) as a function of N for several compounds.

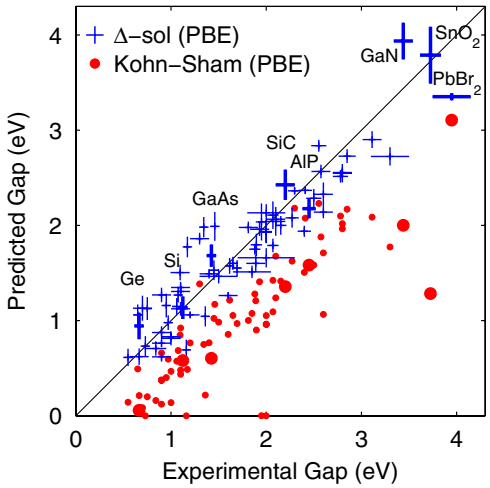

FIG. 4 (color online). E_{FG} (Δ-sol) and E_{KS} using PBE vs E_{Exp} for 74 compounds with only s and p valence electrons. Select compounds are highlighted with bolder symbols.

it lies in the expected range supports our interpretation of the Δ-sol method as based on addition or removal of an electron to a unit defined by the size of the XC hole. This method thus circumvents the delocalization problem in local and semilocal functionals and the murky interpretation of individual Kohn-Sham eigenstates as corresponding to single-particle excitations, and relies instead on the well-proven accuracy of total energies in DFT.

Compounds with only s and p valence electrons are most similar to the HEG and are therefore most amenable to the Δ-sol scheme. Fig. 4 shows the predicted E_{FG} for 74 such compounds for PBE. Significant improvements (mean absolute error $\epsilon = 0.22$ eV and standard deviation $\sigma = 0.17$ eV) are obtained compared to Kohn-Sham gaps ($\epsilon = 0.73$ eV, $\sigma = 0.41$ eV). For half of the compounds the prediction error is less than 0.2 eV, and for 90% it is less than 0.5 eV. The difference between E_{FG} and E_{KS} is not simply correlated with E_{KS}.

We note that these calculations are performed on charged periodic cells with the requisite neutralizing backgrounds and therefore subject to image-charge interaction errors. We calculate the leading monopole term in the image-charge correction [29], $\Delta E_0 = \frac{\alpha n^2}{2\varepsilon L}$, where α is the Madelung constant, ε the bulk (electronic) dielectric constant, and L the size of the cell. We use density functional perturbation theory [30] to calculate ε. We find that $\overline{|\Delta E_0|} = 0.17$ eV for $N^* = 72$ in PBE for the spd test set, and including ΔE_0 gives $\epsilon = 0.40$ eV. As the Madelung term has been shown to overestimate the image-charge error in bulk solids [31], this gives an upper limit estimate of the effects of image-charge errors on our results. We note that care must be taken for materials with small ε or large unit cells, which are associated with larger image-charge errors under the Δ-sol approach.

Despite its simplicity in concept and implementation, Δ-sol performs similarly or better in terms of accuracy to many other state-of-the-art band gap prediction methods. Table II shows band gaps for a set of 12 compounds for which predicted band gaps from the recently-proposed MBJ [13] potential are available. The average errors for both methods in this test set are 0.3 eV. It should be noted that the MBJ potential is fitted to reproduce gaps in a larger

TABLE I. The N^* parameter and accuracy of Δ-sol for band gap prediction in solids, for LDA, PBE, and AM05. The numbers of compounds M for different XC functionals are different because of differences in the availability and reliability of PAW potentials. N^*_{best} is obtained from minimizing the mean absolute error ϵ, and σ is the standard deviation of the absolute errors. For comparison, mean absolute errors and standard deviations are also shown for Kohn-Sham gaps.

Test set	XC	M	N^* min	N^* best	N^* max	Δ-sol ϵ	Δ-sol σ	Kohn-Sham ϵ	Kohn-Sham σ
spd	LDA	109	50	63	80	0.25	0.18	0.93	0.58
spd	PBE	126	59	72	88	0.23	0.16	0.84	0.50
spd	AM05	118	60	76	91	0.23	0.17	0.82	0.50
sp	LDA	57	43	56	78	0.22	0.18	0.74	0.39
sp	PBE	74	52	68	87	0.22	0.17	0.73	0.41
sp	AM05	72	52	70	92	0.22	0.17	0.70	0.38

TABLE II. Comparison of Kohn-Sham (KS), MBJ-Kohn-Sham [13], and Δ-sol gaps for LDA. All figures are in eV.

Compound	E_{Exp}	$\mathrm{KS}_{\mathrm{LDA}}$	$\mathrm{MBJ\text{-}KS}_{\mathrm{LDA}}$	$\Delta\text{-sol}_{\mathrm{LDA}}$
C	5.5	4.1	4.9	5.3
Si	1.1	0.5	1.2	1.0
Ge	0.7	0.0	0.9	0.9
SiC	2.2	1.4	2.3	2.4
BN (cubic)	6.2	4.4	5.9	5.8
GaN	3.4	1.6	2.8	3.9
GaAs	1.4	0.3	1.6	1.5
AlP	2.5	1.5	2.3	2.1
ZnS	3.7	1.8	3.7	3.6
CdS	2.5	0.9	2.7	3.0
AlN	6.1	4.2	5.6	5.3
ZnO	3.3	0.8	2.7	3.5

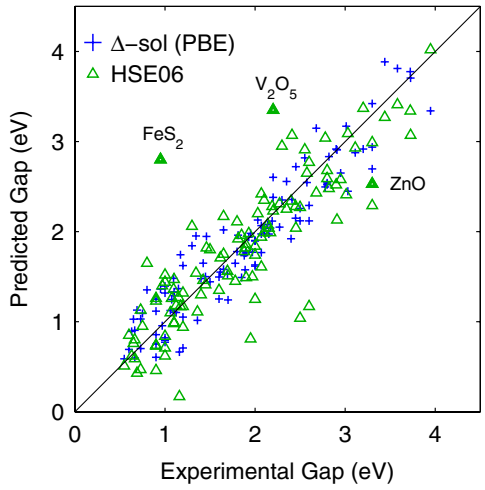

FIG. 5 (color online). A comparison of E_{FG} computed with Δ-sol (PBE) and E_{KS} (HSE06) vs E_{Exp} for 95 compounds.

range (up to 13 eV), whereas N^* is specifically geared towards a narrower range of gaps (0.5 to 4 eV).

We have performed screened-hybrid (HSE06) calculations for 95 compounds in the test set and the results are shown in FIG. 5. HSE06 predicts band gaps for typical semiconductors to better accuracy than Δ-sol ($\epsilon =$ 0.26 eV for HSE06 vs 0.31 eV for Δ-sol for the set in Table II), but has larger errors for transition-metal compounds ($\epsilon = 0.41$ eV for HSE06 vs 0.26 for Δ-sol for the set in FIG. 5). Once again, the parameters in HSE06 are fitted [12] to reproduce gaps in a wider range than Δ-sol.

In summary, we have demonstrated Δ-sol as a viable method for predicting band gaps in solids from total-energy differences. To calculate the Δ-sol gap, one should (i) use experimental lattice constants; (ii) determine the number of valence electrons N_0 in the unit cell; (iii) based on the XC functional used, find N^*_{best}, N^*_{min} and N^*_{max} from Table I; (iv) calculate the energies $E(N_0)$, $E(N_0 + n)$, and $E(N_0 - n)$, where $n = N_0/N^*_{best}$; (v) obtain the gap from $E_{FG} = [E(N_0 + n) + E(N_0 - n) - 2E(N_0)]/n$; and (vi) if uncertainties in E_{FG} are needed, repeat steps 4 and 5 with N^*_{min} and N^*_{max}. This amounts to a few total-energy calculations and can be used with any DFT implementation. Despite the simplicity and efficiency, the accuracy of Δ-sol rivals that of methods that require recoding or are more expensive. The highest accuracies are obtained for compounds with only s and p valence electrons.

This work was supported by Eni S.p.A. under the Eni-MIT Alliance Solar Frontiers Program and Chesonis Family Foundation under the Solar Revolution Project. The authors acknowledge helpful discussions with Patrick Lee, John Joannopoulos, Nicola Marzari, Mark van Schilfgaarde, Troy van Voorhis, and Rickard Armiento. We also thank Anubhav Jain and the Materials Genome project for assistance in automating calculations.

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] D. C. Langreth and J. P. Perdew, Phys. Rev. B **21**, 5469 (1980).
[4] C. S. Wang and W. E. Pickett, Phys. Rev. Lett. **51**, 597 (1983).
[5] O. Madelung, *Semiconductors: Data Handbook* (Springer-Verlag, New York, 2004), 3rd ed.
[6] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).
[7] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Phys. Rev. B **77**, 115123 (2008); P. Mori-Sánchez, A. J. Cohen, and W. Yang, Phys. Rev. Lett. **100**, 146401 (2008).
[8] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **35**, 4170 (1987).
[9] L. Hedin, Phys. Rev. **139**, A796 (1965).
[10] E. Runge and E. K. U. Gross, Phys. Rev. Lett. **52**, 997 (1984).
[11] M. Städele, M. Moukara, J. A. Majewski, P. Vogl, and A. Görling, Phys. Rev. B **59**, 10031 (1999).
[12] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993); C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988); J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996); J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003); A. V. Krukau *et al.*, J. Chem. Phys. **125**, 224106 (2006).
[13] F. Tran and P. Blaha, Phys. Rev. Lett. **102**, 226401 (2009).
[14] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, England, 2004).
[15] J. Langer and S. Vosko, J. Phys. Chem. Solids **12**, 196 (1960).
[16] X. Gonze *et al.*, Z. Kristallogr. **220**, 558 (2005).
[17] G. D. Mahan, *Many Particle Physics* (Kluwer, New York, 2000), 3rd ed.
[18] G. Ortiz and P. Ballone, Phys. Rev. B **50**, 1391 (1994).
[19] M. Ernzerhof and J. P. Perdew, J. Chem. Phys. **109**, 3313 (1998).
[20] A. D. B. W. Kohn and R. G. Parr, J. Phys. Chem. **100**, 12 974 (1996).
[21] J. P. Perdew, K. Burke, and Y. Wang, Phys. Rev. B **54**, 16 533 (1996).
[22] P. Fulde, *Electron Correlations in Molecules and Solids* (Springer, New York, 1995), 3rd ed.
[23] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[24] R. Armiento and A. E. Mattsson, Phys. Rev. B **72**, 085108 (2005); A. E. Mattsson and R. Armiento, Phys. Rev. B **79**, 155101 (2009).
[25] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11 169 (1996); G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[26] P. E. Blöchl, O. Jepsen, and O. K. Andersen, Phys. Rev. B **49**, 16 223 (1994).
[27] L. Wang, T. Maxisch, and G. Ceder, Phys. Rev. B **73**, 195107 (2006).
[28] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, J. Phys. Condens. Matter **9**, 767 (1997).
[29] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).
[30] S. Baroni and R. Resta, Phys. Rev. B **33**, 7017 (1986).
[31] S. Lany and A. Zunger, Phys. Rev. B **78**, 235104 (2008); C. Freysoldt, J. Neugebauer, and C. G. Van de Walle, Phys. Rev. Lett. **102**, 016402 (2009).