General approach for band gap calculation of semiconductors and insulators

Julia Gusakova[**,1], **Beng Kang Tay**[1], and **Vasilii Gusakov**[*,2]

[1] Novitas Center, Nanyang Technological University, 50 Nanyang Avenue, 639798 Singapore, Singapore
[2] Scientific-Practical Materials Research Center of NAS of Belarus, 19 P. Brovki, 220072 Minsk, Belarus

* Corresponding author: e-mail gusakov@ifttp.bas-net.by, Phone: +375 172841290
** e-mail julia001@e.ntu.edu.sg

We have proposed a new approach to the calculation of the band gap of solids. The suggested method does not use any adjustable parameters and is based on the DFT total energy calculation only. The calculated band gaps are in good agreement with experimental ones both for narrow-gap semiconductors (Ge) and insulators with large band gap (Kr). Accurate band gap values were obtained when method was applied to two dimensional material such as MoS_2 monolayer, as well as for bulk form of MoS_2. For investigated materials, the mean absolute error (MAE) of the proposed method is about 0.056 eV, and is significantly lower than the MAE for the currently used most accurate methods such as GW (MAE is 0.54 eV) and HSE (MAE is 0.87 eV).

1 Introduction Density functional theory (DFT) [1] is a tenet for electronic-structure calculations across wide variety of fields, from organic chemistry to condensed matter physics [2]. Although for many properties, DFT produces results which correlate with good accuracy with experimental ones, for electronic properties it is not always the case due to the existence of well-known band gap problem [3]. Widely used LDA and GGA approximations of the exchange-correlation energy functional [4–6] lead to electronic structure and band gap value not accurate enough or underestimated severely.

Various approaches have been developed to overcome this issue and to improve prediction of electronic properties of materials [7]. This includes hybrid functionals such as HSE [8–10] which has been developed and tested on wide groups of semiconductors. Approximate effective potential was introduced by Becke and Johnson which resembles optimized effective potential (OEP) for exchange [11] and later was introduced modified BJ potential (MBJ) [12]. Also have been developed LSDA + U [13] and GW approximation which represents Green's function based method [14].

These approaches improve the description of band gaps of semiconductors comparing with results available from LSDA. Nevertheless, for some materials, the methods mentioned above still give big errors for band gaps (∼20%). The problem is complicated by the lack of correlation between calculated and experimental band gaps. Some methods [12] use adjustable parameters, which is not appropriate for *ab initio* methods. Moreover, it should be mentioned that hybrid functionals and GW methods are very computationally expensive. This fact complicates significantly theoretical search for new (1D–3D) materials with desired electronic (optical, photovoltaic, thermoelectric) properties.

Recently [15], we offered a new general approach for evaluation of a band gap of semiconductors. In the present paper, we present the application of the proposed method, which is based on total energy calculation only, to the wide band gap insulators, van der Waals solids, and two dimensional materials.

2 Theoretical background The main expression for band gap calculation has been derived using total energy functionals of charge density only, which are represented as

Phys. Status Solidi A 213, No. 11 (2016)

Original

Paper

2835

a sum of kinetic energy, interaction with external field, Hartree, and exchange-correlation terms:

$$E = T_S[n(\boldsymbol{r})] + E_{ext}[n(\boldsymbol{r})] + E_H[n(\boldsymbol{r})] + E_{xc}[n(\boldsymbol{r})]. \quad (1)$$

The fundamental band gap is defined as the difference between the first ionization potential (I) and the first electron affinity (A) [16]:

$$\begin{aligned} E_g &= A_e^\infty(N) - A_e^\infty(N+1) \equiv I - A \\ &= E_{N+1}^{(0)} + E_{N-1}^{(0)} - 2E_N^{(0)}, \end{aligned} \quad (2)$$

where $A_e^\infty(N), A_e^\infty(N+1)$ are works on the transfer of one electron from HOMO of a crystal (with N or $N+1$ electrons, respectively) to infinity; $E_{N-1}^{(0)} = E_N^{(0)} + A_e^\infty(N)$, $E_N^{(0)} = E_{N+1}^{(0)} + A_e^\infty(N+1)$, $E_N^{(0)}$ is the total energy of a neutral ($Z = 0$) crystal with N electrons.

In some cases, band gap could be obtained through Kohn–Sham energy eigenvalues:

$$E_g = \varepsilon_{N+1}^{KS}(N+1) - \varepsilon_N^{KS}(N) + C = E_g^{KS} + C, \quad (3)$$

where E_g^{KS} is Kohn–Sham band gap and C is derivative discontinuity of exchange-correlation energy [16]. When the band gap is estimated through Kohn–Sham energies, should be kept in mind that Kohn–Sham eigenvalues are mathematical objects which do not have any known physical meaning. The only exception is the energy of the highest occupied orbital measured relatively to the vacuum level which equals to ionization energy [1].

In our analysis, we expressed the band gap using works to transfer two electrons from HOMO of a crystal (with N or $N+2$ electrons) to infinity: $A_{2e}^\infty(N)$ and $A_{2e}^\infty(N+2)$. In this case, the formula (2) takes the following form:

$$\begin{aligned} E_g &= 0.5\big(A_{2e}^\infty(N) - A_{2e}^\infty(N+2)\big) \\ &= 0.5\big(E_{N+2}^{(0)} + E_{N-2}^{(0)} - 2E_N^{(0)}\big). \end{aligned} \quad (4)$$

On the basis of Eq. (4), we considered two ensembles with total charge equal to zero (see Fig. 1). The first ensemble consists of two neutral supercells, while the second contains charged supercells with total charges $+2$ and -2. For these ensembles, we considered difference between true (exact) total energy functionals and approximated (local density approximation):

$$\Delta_{XC}^\infty = (E_{ens.2} - E_{ens.1})^{(0)} - (E_{ens.2} - E_{ens.1})^{(LDA)}, \quad (5)$$

where $E_{ens.2} - E_{ens.1} = E_{Z=+2} + E_{Z=-2} - 2E_{Z=0}$, E_Z is total energy of the supercell, calculated in the charge state Z and for a given approximation of the exchange-correlation energy. We decomposed exact exchange-correlation energy as sum of the local E_{xc}^{LDA} and nonlocal parts and used Eq. (4) for the ensembles. As a result, we have obtained the nonlinear equation for the exact band gap as a sum of the
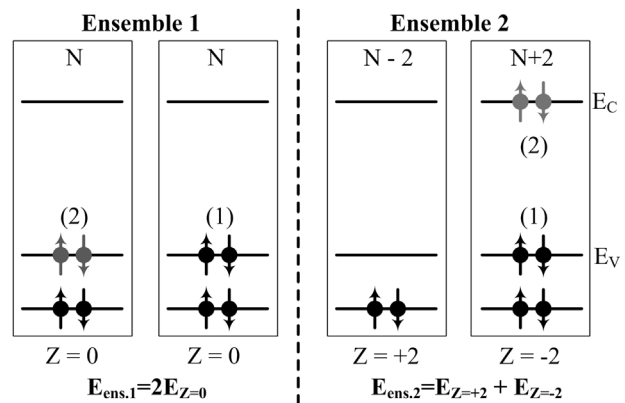
Figure 1 Schematic drawing of ensembles both containing two systems. Each system is a supercell with total charge Z.

band gap, calculated in LDA approximation of the exchange-correlation energy, and two correction terms:

$$E_g^{(0)} = E_g^{(LDA)} + \frac{1}{2}\Delta_{XC}^\infty - \frac{1}{2}E_{(1,0)}^{(\infty,xc)}\left(1 + \exp\left(-\frac{E_g^{(0)}}{E_0}\right)\right), \quad (6)$$

where $E_g^{(LDA)}$ is band gap in LDA approximation calculated using Eq. (4), E_0 is used for the dimensionless energy in the exponent (in the following calculations is taken equal $E_0 = 1\text{eV}$). The correction terms Δ_{XC}^∞ and $E_{(1,0)}^{(\infty,xc)}$ are determined by a nonlocal part of the exchange-correlation energy. $E_{(1,0)}^{(\infty,xc)}$ is a nonlocal part of the exact exchange correlation energy of interaction of the pair electrons (1) (see Fig. 1) with all electrons in lower energy bands and can be represented in the following form:

$$E_{(1,0)}^{(\infty,xc)} = \left(E_{Z=0}^{(0)} - E_{Z=0}^{(LDA)}\right) - \left(E_{Z=2}^{(0)} - E_{Z=2}^{(LDA)}\right). \quad (7)$$

Here we should emphasize the fact that the exact form of the exchange-correlation energy is not available and could only be approximated. So errors, in the calculation of the exact value of the band gap $E_g^{(0)}$ are determined by the precision of calculation Δ_{XC}^∞ and $E_{(1,0)}^{(\infty,xc)}$, and stem from two main approximations, namely the use of pseudopotentials and approximation of the exchange-correlation energy.

The final equation obtained for band gap (6) is based on the DFT total energy calculation only, does not contain any adjustable parameters, and consequently, is less expensive computationally comparing to the GW method (a more detailed derivation of the final equation for the band gap (Eq. (6)) and analysis of terms Δ_{XC}^∞ and $E_{(1,0)}^{(\infty,xc)}$ will be published separately).

3 Results and discussion As it follows from Eq. (6) to calculate $E_g^{(0)}$, we need to calculate the LDA band gap and correction terms Δ_{XC}^∞ and $E_{(1,0)}^{(\infty,xc)}$. The calculation of $E_g^{(LDA)}$ has been performed in a supercell approximation using norm-conserving [17] and PAW [18] pseudopotentials in

Quantum Espresso package [19] and using Eq. (4). There were 216 or 256 atoms in the supercell for the $Fd\bar{3}m$, $F\bar{4}\bar{3}m$, or $P6_3mc$ crystal symmetry, respectively. Geometry of the supercell was optimized by minimizing the total energy as function of the coordinates of all the atoms. Positions of atoms were relaxed until the forces were less than $2 \times 10^{-3}\,\text{eV\,\AA}^{-1}$. For the Kr and Xe fcc crystal structures, the experimental lattice constants were used. As an approximation of the exact expression for the exchange-correlation energy B3LYP [20, 21] approximation was used almost for all materials, while for Xe and Kr PBE [22] approximation was used as B3LYP pseudopotentials were not available. Correction terms $\Delta_{\text{XC}}^{\infty}$ and $E_{(1,0)}^{(\infty,\text{xc})}$ have been calculated using Eqs. (5) and (7) both in cluster and supercell approximations. The cluster was built from the supercell used in LDA calculations and consisted of about 80 atoms. The dangling bonds of the cluster were saturated with hydrogen atoms. In the cluster calculations, the 6-31G basis set was used and the structure of the cluster was optimized.

3.1 Applications for bulk solids In Table 1, detailed results of the band gap computation are represented for bulk semiconductors, insulators, and 2D material. We have purposely considered semiconductor with a narrow band gap (Ge) and dielectric with wide band gap (Kr). As one can see from Table 2, for all investigated materials calculated band gaps are in a good agreement with experimental ones. Correction terms $\Delta_{\text{XC}}^{\infty}$ and $E_{(1,0)}^{(\infty,\text{xc})}$ account for the nonlocal part of the exchange-correlation energy and substantially affect the calculated value of the band gap. As follows from the definition, the $E_{(1,0)}^{(\infty,\text{xc})}$ value should not vary for isostructural semiconductors and this was confirmed for Si, Ge, and C results demonstrated in Table 1. Comparison of the calculated LDA band gap according to Eqs. (2) and (4) (Tables 1 and 2) clearly shows that Eq. (4) gives more accurate results. Really, for germanium Eq. (2) gives band gap equal to 0, while (4) gives LDA band gap value of 0.89 eV.

It is interesting to compare the nonlocal part of exchange-correlation energy $(-0.5E_{(1,0)}^{(\infty,\text{xc})})$ with the electron–electron correlation energy [23] $\varepsilon_{\text{e}-\text{e}} = -(0.019 - 0.0003 r_{\text{s}})$

Table 1 Theoretical band gaps $E_{\text{g}}^{(0)}$ computed with proposed method, LDA_{2e} band gap values and correction terms (all in eV).

solid		$E_{\text{g}}^{(0)}$(Eq. (6))	LDA_{2e}	$0.5\Delta_{\text{XC}}^{\infty}$	$E_{(1,0)}^{(\infty,\text{xc})}$
Si		1.28	0.87	0.65	0.38
Ge		0.78	0.89	0.20	0.43
C		5.39	4.5	1.11	0.43
BN		6.3	5.3	1.2	0.4
Xe		9.82	6.12	0.067	−7.27
Kr		11.74	7.0	2.86	−3.86
MoS_2	bulk	1.23	1.04	0.096	−0.14
MoS_2	1L	1.71	2.06	−0.12	0.37

Table 2 Theoretical band gap values calculated with our method and experimental band gaps (all in eV). The structure of the solid is indicated in parenthesis. For reference, results obtained with other methods are provided (LDA, GW, MBJLDA). Here LDA band gap values are taken from literature.

solid	$E_{\text{g}}^{(0)}$(Eq. (6))	LDA	GW	MBJLDA	HSE	exp.
Ge (A4)	0.78	0.00a	0.95b	0.85a	0.56c	0.74c
Si (A4)	1.28	0.47a	1.41b	1.17a	1.28c	1.17c
C (A4)	5.39	4.11a	6.18b	4.93a	5.49c	5.48c
BN (B3)	6.3	4.39a	7.14b	5.85a	5.98c	6.25c
Xe (A1)	9.82	5.78a	–	8.52a	–	9.8c
Kr (A1)	11.74	6.76a	–	10.83a	–	11.6c

aLDA and MBJLDA from Ref. [12]; bGW Ref. [24]; cexperimental results from Refs. [10, 25].

calculated per electron. For the nonlocal part of correlation energy, r_{s} is of the order of the lattice constant and for covalent semiconductors (ex. Si), we have $-E_{(1,0)}^{(\infty,\text{xc})}/2 \approx -0.215\text{eV}$ and $\varepsilon_{\text{e}-\text{e}} \approx -0.216\text{eV}$. This demonstrates that for covalent semiconductors the nonlocal part of the exchange-correlation energy is determined by the electron–electron correlation energy.

As one can see from Table 2, for all investigated materials band gaps calculated with our method are in a good agreement with experimental ones. Moreover, theoretical band gaps obtained with proposed method have in general smaller deviation from experiment than band gaps obtained using other theoretical methods (GW, HSE, MBJLDA).

We have evaluated mean absolute error (MAE) and mean absolute percentage deviation (MAPD) for the methods presented in Table 2 for the investigated solids. And for our method MAE is 0.055 eV which is the smallest among mean absolute errors given by other methods: 0.548 (GW), 0.866 (HSE), 0.438 (MBJLDA), and 2.455 eV (LDA). Mean absolute percentage deviation for our method is even more impressive and is 2.93%, while other methods have 16.17 (GW), 10.62 (HSE), 7.73 (MBJLDA), and 49.69% (LDA).

3.2 Applications for 2D materials Two dimensional (2D) materials at present include various classes of materials and one of them are Transition Metal Dichalcogenides (TMDC) [26]. TMDCs are layered materials with strong in-plane bonding and week out-of-plane interactions (van der Waals forces). Chemical formula of TMDCs is MX_2, where M stands for transitional metal and X is chalcogen atom [27]. One of the semiconducting representatives of this class is MoS_2. The most interesting property of TMDC materials is the transition from indirect semiconductors in bulk to direct semiconductors in monolayer [28].

Bulk MoS_2 exhibit hexagonal symmetry and correspond to space group $P6_3/mmc$. MoS_2 monolayer consists of one layer of metallic atoms of molybdenum sandwiched between two planes of chalcogen atoms of sulfur.

Phys. Status Solidi A 213, No. 11 (2016)

Original

Paper

2837

Table 3 Fundamental band gaps, theoretical and experimental of bulk and monolayer MoS_2. LDA_{2e} band gap was computed through charge states 0, +2, −2 (Eq. (4)). For comparison experimental and theoretical results from other methods are presented (all in eV).

solid	$E_g^{(0)}$(Eq. (6))	LDA_{2e}	GW	B3LYP	HSE06	exp.
MoS_2 bulk	1.23	1.04	–	1.93[a]	1.46[a]	1.23[b]
MoS_2 1L	1.71	2.06	2.66[c]	–	2.25[c]	1.74[d]

[a]Ref. [30]; [b]Ref. [31]; [c]Ref. [32]; [d]Ref. [28].

The computations for MoS_2 in bulk and monolayer forms were performed using Perdew-Zunger [4] LDA and Perdew–Burke–Ernzerhof [22] GGA approximations of exchange-correlation energy and pseudopotentials. Brillouin zone was sampled using Monkhorst-Pack approach [29] and was used $12 \times 12 \times 3$ mesh for bulk and $10 \times 10 \times 1$ mesh for monolayer. Kinetic energy cutoff was set to 40 Ry.

For bulk and monolayer MoS_2, band gap was computed using the proposed method (Eq. (6)) and through total energies of charged states (see Eq. (2) and results in Table 3). The computation demonstrated that bulk MoS_2 has indirect band gap, while MoS_2 monolayer has direct band gap. Kohn-Sham band gaps obtained from LDA and PBE calculation are 0.79 and 0.93 eV for bulk MoS_2, and 1.87 and 1.62 eV for MoS_2 monolayer. Using our method for both bulk and monolayer MoS_2 band gaps were predicted with good accuracy comparing with experiment (errors less than 1% for bulk and 1.6% for monolayer). From other theoretical papers for bulk MoS_2 band gap was predicted with errors from 19 (HSE06) to 57% (B3LYP). For monolayer MoS_2 errors for theoretical band gap varied from 29 (HES06) to 53% (GW).

The results in Table 3 clearly demonstrate that the band gap values obtained with our method are more precise than GW and HSE and the method can be applied to 2D materials.

4 Summary The application of proposed by us general method of calculating the band gap to the narrow-gap semiconductors (Ge) and insulators with very wide band gap (Kr) has shown that the method predicts the band gap almost with experimental accuracy. We have shown the crucial role of the nonlocal part of the exchange-correlation energy in accurate determination of a band gap of semiconductors and insulators. The method is based on the DFT total energy calculation only and does not introduce any adjustable parameters. As well, computational costs for accurate band gap calculation with proposed method are less computationally expensive than GW. Obtained results also clearly demonstrate applicability of the method for complex materials which have layered structure in bulk with van der Waals interactions between layers (MoS_2) and for 2D layers (MoS_2 monolayer).

References

[1] W. Kohn, Rev. Mod. Phys. **71**, 1253 (1999).
[2] R. Martin, Electronic Structure: Basic Theory and Practical Methods (Cambridge University Press, Cambridge, 2008).
[3] S. Kümmel and L. Kronik, Rev. Mod. Phys. **80**, 3 (2008).
[4] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[5] J. P. Perdew and Y. Wang, Phys. Rev. B **33**, 3399 (1986).
[6] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).
[7] J. P. Perdew and K. Schmidt, AIP Conf. Proc. **577**, 1 (2001).
[8] J. Heyd and G. E. Scuseria, J. Chem. Phys. **120**, 7274 (2004).
[9] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003).
[10] J. Heyd, J. E. Peralta, G. E. Scuseria, and R. L. Martin, J. Chem. Phys. **123**, 174101 (2005).
[11] A. D. Becke and E. R. Johnson, J. Chem. Phys. **124**, 221101 (2006).
[12] F. Tran and P. Blaha, Phys. Rev. Lett. **102**, 226401 (2009).
[13] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).
[14] W. G. Aulbur, L. Jonsson, and J. W. Wilkins, Solid State Phys. **54**, 1 (2000).
[15] J. Gusakova, B. K. Tay, and V. Gusakov, Efficient DFT method for calculating band gap and deep levels in semiconductors, 28th ICDS Conference, Helsinki, Finland, 2015.
[16] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).
[17] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
[18] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[19] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys.: Condens. Matter **21**, 395502 (2009).
[20] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[21] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).
[22] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[23] O. Madelung, Introduction to Solid-State Theory, Springer Ser. Solid-State Sci., Vol. 2 (Springer-Verlag, Berlin, Heidelberg, 1978).
[24] M. Shishkin, M. Marsman, and G. Kresse, Phys. Rev. Lett. **99**, 246403 (2007).
[25] E. E. Koch, V. Saile, N. Schwentner, and M. Skibowski, Phys. Chem. Lett. **28**, 562 (1974).
[26] A. K. Geim and I. V. Grigorieva, Nature **499**, 419 (2013).
[27] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano, Nature Nanotechnol. **7**, 699 (2012).
[28] A. Kuc, N. Zibouche, and T. Heine, Phys. Rev. B **83**, 245213 (2011).
[29] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[30] W. Li, C. F. J. Walther, A. Kuc, and T. Heine, J. Chem. Theory Comput. **9**, 2950 (2013).
[31] K. K. Kam and B. A. Parkinson, J. Phys. Chem. **86**, 463 (1982).
[32] Y. Ding, Y. Wang, J. Ni, L. Shi, S. Shi, and W. Tang, Physica B **406**, 2254 (2011).