Accurate determination of band gaps within density functional formalism

Prashant Singh,1 Manoj K. Harbola,2 Biplab Sanyal,3,4 and Abhijit Mookerjee1

1S N Bose National Centre for Basic Sciences, Salt Lake City, Kolkata 700098, India
2Department of Physics, Indian Institute of Technology, Kanpur 208016, India
3Department of Physics and Astronomy, Uppsala University, Box-516, 75120 Uppsala, Sweden

In this paper, we report an adaptation of the Harbola-Sahni (HS) exchange potential to the tight-binding linear muffin-tin orbital (TB-LMTO) method to determine band gaps (BGs) of solids accurately. We show that the electrostatic basis of derivation of the Harbola-Sahni potential allows this nonvariational approach to improve substantially over local-density approximation derived BGs, bringing them very close to experimental values. That the accuracy of the HS potential is directly responsible for the determination of correct BGs is demonstrated by performing similar calculations with an accurate model potential that too leads to BGs close to their experimental values. Moreover, ground-state properties like equilibrium lattice parameters and bulk moduli (BM) for various semiconductors like C, Si, AlN, AlP, BP, and 3C-SiC calculated with the HS approach are in close agreement with the experiments. The clear physical interpretation of HS potential leads us to suggest exploring its use for calculating various properties of solids.

I. INTRODUCTION

The description of a many-body interacting system of valence electrons in a solid is one of the difficult problems in physics. Perhaps the most successful first-principles approach to the electronic structure of solids is density functional theory (DFT). The ideas behind DFT are quite simple and remarkably easy to implement for numerical studies. However, the inaccuracy in predicting the band gaps (BGs) of semiconductors and insulators has been one of the recurring problems in this approach. The crux of the matter lies in the setting up of the auxiliary, single-particle Kohn-Sham equation. In the traditional DFT, the auxiliary Hamiltonian is obtained variationally. As a result, DFT is applicable only to the ground state. It is well understood that the spectrum and orbitals of the Kohn-Sham equation have no specific significance beyond the fact that they are used to obtain the ground-state density and the kinetic-energy contribution. The exception to this is the highest occupied orbital whose eigenenergy is equal to the negative of the ionization potential I of the system. Thus the eigenenergy corresponding to the top of the valence band is −I. This gives hope that the band gap \( E_g \) of a semiconductor will be equal to the Kohn-Sham gap \( E_{KS}^{KS} \). This is based on the fact that the addition of one electron to a bulk system will not change its density by any significant amount hence it is expected that all its eigenvalues will remain essentially unchanged. Thus the eigenenergy corresponding to the bottom of the conduction band should be equal to −A, where A is the electron affinity, therefore \( E_{KS}^{KS} \) will be equal to the fundamental gap \( I - A \). However, as is well known, this does not turn out to be the case.

The reason for the difference between the Kohn-Sham gap and the fundamental gap is the discontinuous change by a constant \( \Delta_{xc} \) in the Kohn-Sham potential when the electron number changes by a fractional amount across an integer. For example, in the helium ion He\(^+\), as the electron number is increased fractionally by \( \delta \) from 1 to \( 1 + \delta \), the exchange-correlation potential jumps by 1.1 atomic units.8

Thus

\[
E_g = E_{KS}^{KS} + \Delta_{xc}.
\]  

The discontinuity could be as large as 50% or even more of the true gap. For example, in silicon the \( E_{KS}^{KS} \) calculated with the local-density approximation (LDA) is 0.49 eV whereas the experimental gap is 1.17 eV. This would mean that \( \Delta_{xc} \) is 0.68 eV if the LDA gap is close to the true Kohn-Sham gap. The latter, however, is not known. Indeed, at present, no straightforward method exists to estimate the exact value of the discontinuity \( \Delta_{xc} \).

There have been considerable efforts to improve the BGs within the framework of Kohn-Sham DFT. Extending the idea of optimized potential method (OPM), first proposed by Talman and Shadwick for atoms to solids, Kotani combined exact exchange (EXX) with LDA correlation to obtain BGs in good agreement with the experiments. Presently, within first-principle approaches for extended systems, the GW approximation explains electronic band properties most accurately. However, the computational efforts required for the GW method are much heavier, thus limiting its application to relatively smaller systems. In this paper, we propose an alternative approach to obtain the exchange potential, which is calculated as the work done in moving an electron in the electric field produced by its Fermi-Coulomb hole. This physical interpretation of the Kohn-Sham potential was provided by Harbola and Sahni (HS). Using the HS approach within the exchange-only approximation, where the potential is evaluated from the Fermi hole, produces ground-state as well as excited-state properties comparable to EXX and GW but with considerably reduced numerical efforts. The above physical picture and the vastly reduced computational effort make this approach worth following. With this in mind, we use the HS approach to calculate various properties of materials within TB-LMTO in the atomic sphere approximation (ASA). We find that the present method gives all properties including the BGs in good agreement with experiments. To
understand why methods like EXX or the HS give BGs close to experiments, we also perform calculations by employing an accurate model potential given by the van Leeuwen and Baerends (LB)\textsuperscript{14} correction to the LDA exchange-correlation potential, and show that the accuracy of potential in the outer region of AS leads to correct BGs.

II. METHODOLOGY

A. Harbola-Sahni potential

In the post-Born-Oppenheimer many-body Hamiltonian of the interacting valence electron system, the contribution to the total energy by the electron-electron interaction terms is

\[ E_{ee} = \frac{1}{2} \int \int \frac{\rho(r)\rho(r')}{|r-r'|} dr dr'. \]

The joint probability density can be written as

\[ \rho_{ee}(r, r') = \rho(r)\rho(r')[1 - C(r, r')] \]

where \( C(r, r') \) is the correlation function. The electron-electron part of the total energy becomes

\[ E_{ee} = \frac{1}{2} \int \int \frac{\rho(r)\rho(r')}{|r-r'|} dr dr' + \frac{1}{2} \int \int \frac{\rho(r)\rho_{ee}(r, r')}{|r-r'|} dr dr', \]

where \( \rho_{ee}(r, r') = -\rho(r)c(r, r') \) is the Fermi-Coulomb hole charge distribution. The physical interpretation of the Fermi-Coulomb hole is the deficit in the density of electrons at \( r' \) caused by the presence of an electron at \( r \).

We can approach the problem of defining the exchange part of the Kohn-Sham potential by its electrostatic definition as the work done to move an electron in the electric field \( \mathcal{E}(r) \) produced by its Fermi hole \( \rho_i(r, r') \). Thus the exchange potential is given as

\[ W_{HS}(r) = - \int_{-\infty}^{r} \mathcal{E}(r') \cdot d\ell, \]

where

\[ \mathcal{E}(r) = \int \frac{\rho_i(r, r')}{|r-r'|^3} (r - r') d\ell'. \]

For a set of single-particle orbitals \( \{\phi_{i\sigma}\} \) with occupation \( f_{i\sigma} \) for the \( i \)th orbital with spin \( \sigma \), the Fermi hole is given\textsuperscript{15} as

\[ \rho_i(r, r') = \frac{1}{\rho(r)} \sum_{\sigma} \sum_{i,j} f_{i\sigma} f_{j\sigma}^* \phi_{i\sigma}^*(r) \phi_{i\sigma}^*(r') \phi_{j\sigma}(r') \phi_{j\sigma}(r). \]

The virial theorem\textsuperscript{22} gives the exchange energy from the HS potential \( W_{HS}(r) \) through the relationship

\[ E_x = - \int \rho(r) r \cdot \nabla W_{HS}(r) dr. \]

There have been several applications of these ideas to estimate accurately the ground-state energies as well as the energies of excited states of several atomic systems.\textsuperscript{23-27} It is well known that the latter are notoriously difficult to estimate theoretically. But the excellent results reported for both the ground and excited states encouraged us to apply this technique to solids. We mention that the HS approach for the exchange-only calculation is intimately related to the EXX method based on OPM as discussed briefly later in this section. For the details of their connection, we refer the reader to Refs. 2, 7, 8, and 12.

Using the approach outlined above, the Kohn-Sham equation is

\[ \left\{ -\frac{1}{2} \nabla^2 + V_{\text{eff}}(r) \right\} \phi_{i\sigma}(r) = \epsilon_{i\sigma} \phi_{i\sigma}(r) \]

The aim of this work is to implement Eq. (7) in TB-LMTO-ASA\textsuperscript{13} to calculate various properties of solids. Within TB-LMTO-ASA, the lattice space is divided into ion-core centric atomic spheres (ASs) with overlap \( < 10\% \), where ASA demands that the sum of AS volumes equals the cell volume. For open lattices this is not possible so the remaining volume is filled with empty spheres (ESs) which do not have associated ion cores but have charge associated with them. Inside the AS, the Kohn-Sham potential is obtained by using only the exchange potential given by Eq. (3). On the other hand, in the ES it is calculated with the standard LDA including both the exchange and correlation potentials. As pointed out above, within the ASA Eq. (3) gives a well defined exchange potential since the underlying field is curl free. The potential is fixed to be \(-1/r \) a.u. at the AS boundary. Correlation is included in the ES because the density in this region is very small and therefore the effect of correlation is expected to be as significant as that of the exchange. Application of the LDA in the ES is justified as the absence of a core here makes the electron gas reasonably homogeneous. Using these potentials, Eq. (7) is then solved self-consistently. The results thus obtained are given in Table I under the column HS-EX and are discussed in Sec. III.

As pointed out in the beginning, the KS gap differs from the true fundamental gap by the integer derivative discontinuity in the exchange-correlation potential. However, it has been shown that for finite systems, the accurate treatment of exchange-correlation potential, particularly in the outer regions, leads\textsuperscript{28} to eigenvalue differences between the highest occupied orbital and unoccupied orbitals energies close to true excitation energies. This indicates that the same may hold true for solids too.

B. Comparison of the HS approach with the OPM

As mentioned in the previous section, the HS potential can also be derived directly from the Schrödinger equation. However, it is rendered approximate because of the neglect of difference in the true kinetic energy and the Kohn-Sham kinetic
TABLE I. The BGs from the LMTO-HS-EX compared with the LMTO-LDA, GW (Ref. 32), LMTO-HSP (the HS potential (HSP) is used both inside and outside the AS), LMTO-EXX (results for diamond and silicon taken from Kotani (Ref. 10) and results for AlN and 3C-SiC are taken from Görling et al. (Ref. 29), LMTO-LB and experiments (Refs. 35–41). The HS-EX, the HSP, the LDA, and the LB are evaluated within TB-LMTO-ASA by us. C and Si are in diamond structure while AlN, AlP, BP, and SiC are in zinc-blende (ZB) structure.

<table>
<thead>
<tr>
<th>Element</th>
<th>HS-EX</th>
<th>LDA</th>
<th>GW</th>
<th>HSP</th>
<th>EXX</th>
<th>LB</th>
<th>Expt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>5.47</td>
<td>2.70</td>
<td>6.03</td>
<td>5.51</td>
<td>5.12</td>
<td>5.18</td>
<td>5.48</td>
</tr>
<tr>
<td>Si</td>
<td>1.24</td>
<td>0.49</td>
<td>1.37</td>
<td>1.86</td>
<td>1.93</td>
<td>1.21</td>
<td>1.17</td>
</tr>
<tr>
<td>AlN</td>
<td>5.05</td>
<td>2.44</td>
<td>4.90</td>
<td>4.91</td>
<td>5.03</td>
<td>5.13</td>
<td>5.11</td>
</tr>
<tr>
<td>AlP</td>
<td>2.53</td>
<td>1.16</td>
<td>2.86</td>
<td>2.12</td>
<td>2.75</td>
<td>2.51</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>2.22</td>
<td>1.51</td>
<td>1.90</td>
<td>2.54</td>
<td>2.09</td>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td>3C-SiC</td>
<td>2.88</td>
<td>1.38</td>
<td>2.76</td>
<td>3.01</td>
<td>2.52</td>
<td>2.58</td>
<td>2.42</td>
</tr>
</tbody>
</table>

III. RESULTS AND DISCUSSION

A. Calculations with Harbola-Sahni potential

In Fig. 1, we have plotted the potential for Si in the AS and the ES. There is a small discontinuity in the potential at the AS boundary. This is an artifact of ASA. Within the AS, the LDA is relatively less negative in comparison to the HS-EX to exchange due to noncancellation of self-interaction. In TB-LMTO the AS and ES orbitals have fractional charges which have been taken care of while calculating the radial and Fermi hole charge densities.15 Total valence electrons in the Si-AS using the HS-EX is 3.35, while the LDA has only 3.21 electrons.

For band-structure calculations, we started with Si in the diamond structure, which consists of two interpenetrating fcc lattices with Si atoms at (0,0,0) and (1/4,1/4,1/4) respectively in the unit cell. We have added two other fcc lattices of empty spheres to obtain a close-packed structure within TB-LMTO-ASA. The average Wigner-Seitz radius is set to 2.53 Å for

![Image](313x108 to 558x324)

FIG. 1. (Color online) Anticlockwise, radial density (RD) of the HS-EX potential (top left), the HS-EX (blue line), and the LDA (green line) potentials for the Si-AS (at bottom left) and the ES (at right). Wigner-Seitz radius is kept fixed for both Si-AS and ES at 2.53 Å.
FIG. 2. (Color online) Electronic band structure and density of states of Si calculated within the HS-EX (blue line) and the LDA (green line).

both Si-AS and ES. The basis set used for Si-AS is $(3s^3p^3d)$ and $(1s^2p^3d)$ for ES, which are complete under all symmetry operations and no additional basis atom has been introduced. The tetrahedron integration is performed over the full Brillouin zone with 29 irreducible $k$ points derived from $512 (8 \times 8 \times 8)$ $k$ points. The band structure of Si thus obtained is shown in Fig. 2. A comparison of the dispersion curves obtained with the HS-EX and the LDA shows that the curvature of valence bands are hardly affected in the HS-EX. However, a definite shift has occurred in valence bands while conduction bands shifted with some distortion which is not a rigid shift like the scissor operator method. The energy-dependent distortion invalidates the use of the scissors operator which empirically inserts a rigid shift. The Fermi level has been set at zero. The conduction-band minimum in silicon occurs at about 0.85% of the way to the zone boundary from $\Gamma$ to $X$. The difference between the conduction-band minimum and the valence-band maximum results into a correct indirect $\Gamma$-$X$ gap. The BG of Si produced by the HS-EX is 1.24 eV whereas LDA highly underestimates with a value of 0.49 eV. The values of 1.37 and 1.93 eV obtained within GW$^{32}$ and EXX$^{10}$ respectively are overestimated. For silicon, within the HS-EX, the calculated effective electron mass ($m_e^*/m_0$) is 0.99 while Kittel et al. reported a value of 0.97 ± 0.02. Our calculated effective hole mass ($m_h^*/m_0$) is 0.44 while the experiments by Dexter et al. showed an average hole mass to be 0.39, where $m_0$ is the free-electron rest mass.

Table I shows that our predictions for BGs calculated using the HS-EX are in excellent agreement with experiments. The absence of self-interaction in the HS-EX, contrary to the LDA, localizes the valence bands comparatively more than the conduction bands due to a large exchange contribution by the former. Consequently, instead of subsiding, the HS-EX improves the BGs, demonstrating the applicability of Harbola-Sahni construct to the description of excited states. For completeness, we also show in Table I the results of BGs using pure HS-EX potential (HSP) both in the AS and the ES. The HSP potential is calculated by fixing its value to be $-1/r$ at the boundaries of the AS and the ES. The results obtained by employing the HSP are similar to the HS-EX and the EXX results. However, HE-EX results are closer to experiments because of the inclusion of correlation in the ES.

Next, we have applied the HS-EX based TB-LMTO method to calculate structural properties, e.g., equilibrium lattice parameters (ELPs) and BM of semiconductors as well as metals which have been summarized in Table II. A good agreement is obtained between experimental and calculated ELPs. However, the average deviation of ELPs from the experiments is less than 1% for all systems except the metals like Li and Na which differ by as much as 5%.

### Table II: BM of semiconductors and metals evaluated by fitting the data to the Murnaghan equation of state (Ref. 42).

<table>
<thead>
<tr>
<th>Element</th>
<th>HS-EX</th>
<th>LDA</th>
<th>Expt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a (Å)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>3.60</td>
<td>3.53</td>
<td>3.57</td>
</tr>
<tr>
<td>Si</td>
<td>5.51</td>
<td>5.38</td>
<td>5.43</td>
</tr>
<tr>
<td>AlN</td>
<td>3.84</td>
<td>4.31</td>
<td>4.36</td>
</tr>
<tr>
<td>AIP</td>
<td>5.50</td>
<td>5.41</td>
<td>5.46</td>
</tr>
<tr>
<td>BP</td>
<td>4.60</td>
<td>4.51</td>
<td>4.54</td>
</tr>
<tr>
<td>3C-SiC</td>
<td>4.41</td>
<td>4.30</td>
<td>4.36</td>
</tr>
<tr>
<td>Li</td>
<td>3.27</td>
<td>3.40</td>
<td>3.45</td>
</tr>
<tr>
<td>Na</td>
<td>4.03</td>
<td>4.04</td>
<td>4.21</td>
</tr>
<tr>
<td>Al</td>
<td>3.98</td>
<td>3.99</td>
<td>4.02</td>
</tr>
<tr>
<td>BM (GPa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS-EX</td>
<td>429</td>
<td>457</td>
<td>438</td>
</tr>
<tr>
<td>LDA</td>
<td>85.4</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>HF</td>
<td>208</td>
<td>206</td>
<td>254</td>
</tr>
<tr>
<td>Expt.</td>
<td>99</td>
<td>95</td>
<td>86</td>
</tr>
</tbody>
</table>

B. Calculations with van Leeuwen and Baerends (LB) correction to the LDA potential

As we have seen above, the use of the HS potential in the AS combined with the LDA potential in the ES leads to reasonably accurate BGs for the systems reported in Table I. This is similar to the results obtained with the OPM/EXX and self-interaction corrected LDA (SIC-LDA). This we believe is related to the proper treatment of the exchange-correlation potential in the outer regions of the AS. We understand that the improvement in the values of BGs corresponds to the integer discontinuity of exchange-correlation potential; however, the correctness of the uppermost eigenvalue, the appropriate asymptotic behavior, and integer discontinuity of the exchange-correlation potential are all inter-related. Thus, it is not a coincidence that for a large
The potential has been employed in the past to study the behavior of the model potential in the outer regions of finite systems. Further, the EXX has been shown to possess derivative discontinuity. We now demonstrate that improving the LDA for its behavior in the outer regions of the AS indeed leads to significant improvement of the values of BGs. We do this by calculating the BGs of the systems studied in this paper with a model potential that corrects the LDA in the outer regions of finite systems. The correction to the LDA exchange has been made both in the AS and the ES. The method is easy to implement and computationally less expensive and should allow one to treat complex systems with this method. Also using a model potential, we have shown that improving the potential in outer regions of AS leads to good band gaps and as such we conclude that accurate band gaps can be obtained in such situations. Of course, this method is van Leeuwen and Baerends correction to the LDA exchange-correlation potential and the parameter β = 0.05 and x = |∇ρ|/ρ/1/3.

The correction V_{LB} is motivated by the Becke formula for exchange energy and therefore leads to correct −1/r behavior of the model potential in the outer regions of finite systems. The potential has been employed in the past to study the effect of the correct asymptotic behavior of the potential on response properties of atoms. In this paper, we employ it to obtain the exchange correlation potential in the AS and use it in combination with the LDA potential in the ES to perform self-consistent calculation within TB-LMTO-ASA. The results of the BGs obtained from these calculations are given in Table I. As is evident from these results, the model potential of Eq. (11) gives BGs that are a vast improvement over the LDA results and are in good agreement with experimental values. This shows that an improved exchange-correlation potential, which is particularly accurate in the outer regions of the AS, gives significantly improved results over the LDA. In this connection we further note that modifying the LDA potential by the Becke-Johnson potential also leads to substantially improved BGs for a variety of solids.

IV. CONCLUSION

To conclude, we have developed a first-principles method based on Harbola-Sahni exchange potential within TB-LMTO basis to calculate accurately the band gaps of semiconductors. Not only the calculated band gaps, but also properties like equilibrium lattice parameters and bulk moduli of considered semiconductors are in good agreement with experiments. The method is easy to implement and computationally less expensive and should allow one to treat complex systems with this method. Also using a model potential, we have shown that improving the potential in outer regions of AS leads to good band gaps and as such we conclude that accurate band gaps predicted by HS potential are due to the latter’s accuracy in these regions. However, there are still issues to be settled. We have estimated the local potential due to the Fermi hole caused by the Pauli principle. However, the Coulomb hole, in which even electrons with the different spins cannot come near each other, needs to be incorporated. This will be addressed in the future.

ACKNOWLEDGMENTS

This work was a part of the HYDRA collaboration. P.S. would like to thank O. Eriksson for the kind hospitality during his visit to Uppsala. We would also like to thank O. K. Andersen for the use of the TB-LMTO codes.
38 O. Ambacher (private communication).