Simple correction to bandgap problems in IV and III–V semiconductors: an improved, local first-principles density functional theory

Sujoy Datta
*University of Calcutta and Lady Brabourne College*

Prashant Singh
*Ames Laboratory, psingh84@ameslab.gov*

Chhanda B. Chaudhuri
*Lady Brabourne College*

Debnarayan Jana
*University of Calcutta*

Manoj K. Harbola
*Indian Institute of Technology, Kanpur*

*See next page for additional authors*

Follow this and additional works at: [https://lib.dr.iastate.edu/mse_pubs](https://lib.dr.iastate.edu/mse_pubs)

芡 Part of the Electronic Devices and Semiconductor Manufacturing Commons, and the Semiconductor and Optical Materials Commons

The complete bibliographic information for this item can be found at [https://lib.dr.iastate.edu/mse_pubs/301](https://lib.dr.iastate.edu/mse_pubs/301). For information on how to cite this item, please visit [http://lib.dr.iastate.edu/howtocite.html](http://lib.dr.iastate.edu/howtocite.html).

This Article is brought to you for free and open access by the Materials Science and Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Materials Science and Engineering Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Simple correction to bandgap problems in IV and III–V semiconductors: an improved, local first-principles density functional theory

Abstract
We report results from a fast, efficient, and first-principles full-potential $N$th-order muffin-tin orbital (FP-NMTO) method combined with van Leeuwen–Baerends correction to local density exchange-correlation potential. We show that more complete and compact basis set is critical in improving the electronic and structural properties. We exemplify the self-consistent FP-NMTO calculations on group IV and III–V semiconductors. Notably, predicted bandgaps, lattice constants, and bulk moduli are in good agreement with experiments (e.g. we find for Ge 0.86 eV, 5.57 angstrom, 75 GPa versus measured 0.74 eV, 5.66 angstrom, 77.2 GPa). We also showcase its application to the electronic properties of 2-dimensional $h$-BN and $h$-SiC, again finding good agreement with experiments.

Keywords
DFT, semiconductor, band-gap, electronic structure

Disciplines
Electronic Devices and Semiconductor Manufacturing | Materials Science and Engineering | Semiconductor and Optical Materials

Comments
This is a peer-reviewed, un-copyedited version of an article accepted for publication/published in Journal of Physics: Condensed Matter. IOP Publishing Ltd is not responsible for any errors or omissions in this version of the manuscript or any version derived from it. The Version of Record is available online at 10.1088/1361-648X/ab34ad. Posted with permission.

Authors
Sujoy Datta, Prashant Singh, Chhanda B. Chaudhuri, Debnarayan Jana, Manoj K. Harbola, Duane D. Johnson, and Abhijit Mookerjee

This article is available at Iowa State University Digital Repository: https://lib.dr.iastate.edu/mse_pubs/301
Simple correction to bandgap in IV and III-V semiconductors: an improved first-principles local density functional theory

Sujoy Datta,1,2 Prashant Singh,3,* Chhanda B. Chaudhuri,2 Debnarayan Jana,1 Manoj K. Harbola,4 Duane D. Johnson,3,5,† and Abhijit Mookerjee,6

1Department of Physics, University of Calcutta, Kolkata 700009, India
2Department of Physics, Lady Brabourne College, Kolkata 700017, India
3Ames Laboratory, U.S. Department of Energy, Iowa State University, Ames, Iowa 50011 USA
4Department of Physics, Indian Institute of Technology, Kanpur, 208016, India
5Department of Materials Science & Engineering, Iowa State University, Ames, Iowa 50011, USA
6S. N. Bose National Centre for Basic Sciences, Salt Lake City, Kolkata 700098, India

(Dated: July 18, 2019)

We report results from a fast, efficient, and first-principles full-potential Nth-order muffin-tin orbital (FP-NMTO) method combined with van Leeuwen-Baerends correction to local density exchange-correlation potential. We show that more complete and compact basis set is critical in improving the electronic and structural properties. We exemplify the self-consistent FP-NMTO calculations on group IV and III-V semiconductors. Notably, predicted bandgaps, lattice constants, and bulk moduli are in good agreement with experiments (e.g., we find for Ge 0.86 eV, 5.57 Å, 75 GPa vs. measured 0.74 eV, 5.66 Å, 77.2 GPa). We also showcase its application to the electronic properties of 2-dimensional h-BN and h-SiC, again finding good agreement with experiments.

I. INTRODUCTION

Semiconducting materials remain of great interest due to their central role in modern electronics.1–7 In this era of designing optimal materials, it has become essential to estimate quickly and accurately their bandgaps. Experimentally we would benefit from an efficient computational tool for predicting bandgaps (among many other properties) prior to synthesis. We consider the problem of predicting bandgaps to be “solved” by a method that delivers an accurate result for compounds spanning the whole periodic table and is simultaneously practical to compute. Standard exchange correlation (XC) functionals to density functional theory (DFT), such as local density (LDA) or semi-local generalized gradient (GGA)8,9 approximations, have been the backbone of DFT calculations for decades. Unfortunately, the unphysical self-Coulomb repulsion10 leads to a systematic underestimation of bandgaps.11–13

To characterize the electronic properties of a semiconductor or insulator, the fundamental bandgap ($E_{\text{gap}}$) is a key quantity. $E_{\text{gap}}$ is defined as the difference of ionization energy ($I$) and electron affinity ($A$) of the $N$ electron system. While the Kohn-Sham (KS) gap ($E_{\text{KS}}^{N \rightarrow N+1}$) of non-interacting ($N$) electron system is defined as the difference of the highest occupied (HO=$-I$) and the lowest unoccupied (LU=$-A$) band energies. $E_{\text{KS}}^{N \rightarrow N+1}$ would be same as fundamental gap if $I$ and $A$ are exact; however, KS-DFT using LDA or GGA leads to a large underestimation of $E_{\text{KS}}^{N \rightarrow N+1}$.8,9,15,16 Thus, the fundamental gap is defined as a discontinuous change ($\Delta_{xc}$) in the KS potential: $E_{\text{gap}} = E_{\text{KS}}^{N \rightarrow N+1} + \Delta_{xc}$.17 The $\Delta_{xc}$ in DFT is refer to as the jump discontinuity in the derivative of the total energy with respect to $N$.18,19 Although the LDA exchange energy has a derivative discontinuity at integer $N$, the potential shows no such discontinuity.20 For LDA or GGA8,9 the discontinuity can 50% or larger and no straightforward method exists to estimate $\Delta_{xc}$.21 Due to the $\Delta_{xc}$ contribution to $E_{\text{gap}}$ in KS-DFT, there is pressing need for “improved” functionals to assess quickly the HOMO-LUMO from KS orbitals that is closer to observed bandgaps. To access to the “exact” KS-DFT functional (i.e., not LDA/GGA), the proposed functional should reproduce the $\Delta_{xc}$ in the KS potential.22

Several improved semi-local XC functionals have been constructed using increasingly complex density dependence, but satisfy some exact physical features.23,24 If an approximate (semi-local) functional can reproduce the exact asymptotic behavior in potential, then it is good enough for extracting excited state information.25–27 Some semi-empirical functionals are derived as correction to LDA exchange,20,21,27–35 incorporating density gradients to yield asymptotically well-behaved potentials,21,34–37 which are easily handled mathematically and yield good results with little cost computationally.28,35,37 Yet, care is needed in getting discontinuities in the solid-state limit.38 Similarly, accurate, variational exact-exchange functionals approaching the van Leeuwen-Baerends (vLB) corrected LDA form have appeared,39 but are unexplored for solids.

Here we use the vLB-corrected LDA (LDA+vLB) in full-potential Nth-order muffin-tin orbital (FP-NMTO) method. The self-consistent FP-NMTO uses compact basis set to describe the entire system including the interstitial, and it is accurate whether or not it is close-packed. The error arising when making the energy-independent basis set is minimized in FP-NMTO, which is in contrast to Tight-Binding Linearized MTO (TB-LMTO-ASA).40–44 The FP-NMTO utilizes the interpolation of higher-order energy derivatives of Taylor’s expansion of basis set45–48 and provides a smaller and energy-independent basis set. The FP-NMTO basis set is comparable to localized Wannier functions – a prerequisite
II. COMPUTATIONAL METHOD

We apply self-consistent FP-NMTO within LDA+\textit{v}LB to calculate electronic properties of group IV and III-V semiconductors, and discuss Ge and GaAs. We compare FP-NMTO bandgaps using LDA,\textsuperscript{49} PBE,\textsuperscript{5}\textsuperscript{8} and LDA+\textit{v}LB\textsuperscript{25} with experiments and other existing theories. In FP-NMTO, we use third-order correction to the energy derivatives of wave functions (\(N = 3\)) to make the energy-independent basis set. We use \(8 \times 8 \times 8\) k-point mesh (29 irreducible points) for Brillouin zone sampling. The self-consistent potential converged to a difference of \(10^{-6}\) after several tens of iterations. For lattice constants, we utilized a least square fit of our data to Murnaghan’s equation of state.\textsuperscript{50} The FP-NMTO LDA+\textit{v}LB equilibrium lattice constants are in good agreement with the experiments.\textsuperscript{51,52} All calculations are done self-consistently and non-relativistically, using Andersen mixing scheme to facilitate convergence. The tetrahedron method was used for k-space integration. (See appendix for more description of TB-LMTO-ASA.)

III. THEORETICAL DETAILS

Partial-wave solutions: NMTO vs LMTO

In the TB-LMTO-ASA, atomic spheres are used to describe both the variation of electronic charge and the potential. These spheres need to be space filling with minimal (optimal) overlap (\(\leq 15\%\)). Overlap is essential for atomic bonding, whereas too large an overlap of charge may lead to its over-counting. The potential for a given atom is reasonably long ranged but the atomic charge density is more localized around the ion cores. However, NMTO has two different classes of concentric spheres: potential (MT/ASA) spheres of radius \(s\) and charge spheres (screening/hard spheres) of radius \(a\), whereas \(a < s\), see Fig.1(b). However, the overlap of potential spheres in NMTO can be as large as 50\% so full potential approach is a pre-requisite both in atomic spheres and interstitials.

The solutions of radial KS equations [Eq. 4] for atomic-potential \(\phi^a\) is continued from the core of atom towards the muffin-tin surfaces at \(r = s\). This is matched with the solution for a flat part of potential (as in LMTO, this flat potential is taken as zero potential) \(\phi^{int}\), which is denoted by \(\phi^0\). The \(\phi^0\) is valid up to the boundary of hard sphere surface \((r = a)\). We normalize at the boundary of hard sphere surface \((a)\) such that \(\phi^0|_{a} = 1\).

The normalized wave-functions have superscript “a”:

\[
\begin{align*}
\varphi^a_{Rl}(\epsilon, r) &= \frac{\varphi_{Rl}(\epsilon, r)}{\varphi^0_{Rl}(\epsilon, ar)} \\
\varphi^0_{Rl}(\epsilon, r) &= \frac{\varphi^0_{Rl}(\epsilon, r)}{\varphi_{Rl}(\epsilon, ar)}
\end{align*}
\]  

(1)

The \(\varphi^0_{Rl}(\epsilon, r)\) is matched with screened spherical waves (SSW: \(\psi^a\)) continuously but with a kink at \(a\) and truncated in the region \(ar < r < sR\). \(\varphi_{Rl}(\epsilon, r)\) is truncated outside \(0 < r < sR\). The SSWs grow outswards, which are sorted in two groups depending on the values of \((R,l,m)\): active channels (subscript \(A \equiv RL\)) and passive channels (subscript \(I \equiv RL\)). The matching described above is for the active channels which are truncated inside hard spheres with kinks. The passive channels are substituted smoothly by \(\phi^{as}\) inside hard spheres. Finally, a basis set with elements termed as Kinked Partial Waves (KPW), which are defined in all space, is formed as:

\[
\begin{align*}
\phi^a_{Rlm}(\epsilon, r) &= [\varphi^a_{Rl}(\epsilon, r)R - \varphi^0_{Rl}(\epsilon, rR)]Y_{l,m}(\hat{r}R) \\
&+ \psi^0_{Rlm}(\epsilon, rR)
\end{align*}
\]  

(2)

A linear combination \(\sum_{RL\subset A} \phi^a_{Rlm}(\epsilon, rR)c^a_{Rlm,i}\) of KPWs is a solution of KS equation but not the individual KPWs. A kink cancellation condition (similar to the KKR Green’s function zero-determinant condition) for solution can be achieved such that a kink in any hard sphere is cancelled by the sum of kinks from tails of KPWs from neighboring sites.
This basis set spanned by the KPWs is explicitly energy dependent. Energy-dependent basis set slows down the calculation, which was the problem in exact MTO (EMTO) formalism. In LMTO, energy dependence is removed by a self-consistent choice of energy \( \epsilon_r \) about which a Tailor series expansion of basis functions are performed. The series is truncated after first-order term leading to an error \( \propto (\epsilon - \epsilon_r)^2 \) in the solution of KS equation. The energy-independent set of basis functions, the so-called NMTO are the superposition of KPWs at a mesh of energies (calculated self-consistently) using Newton’s interpolation method such that:

\[
\chi^{0 \ldots N} = \sum_{n=0}^{N} |\phi(\epsilon_n)| L_n^{0 \ldots N}
\]

\[
= |\phi(0)| + |\phi(0)| (E^{0 \ldots N} - \epsilon_0) + \ldots + |\phi(0 \ldots N)| (E^{(N-1) - \epsilon_{N-1}} (E^{0 \ldots N} - \epsilon_0)
\]

where \( \phi[0] \equiv \phi(\epsilon_0); \phi[0 \ldots n] \equiv \phi(\epsilon_0) - \phi(\epsilon_1) + \epsilon_0 - \epsilon_1 \) are the terms in divided difference table. The error in the solution of KS equation becomes

\[
\psi(r) - \psi(\epsilon, r) \propto (\epsilon - \epsilon_0)(\epsilon - \epsilon_1) \ldots (\epsilon - \epsilon_n).
\]

van Leeuwen-Baerends Correction to Exchange

From the Hohenberg-Kohn theorem, the Schrödinger’s equation for many-body system reduces to the KS equation for a system of non-interacting electrons moving in an effective potential due to all other electrons and ions:

\[
T[\rho(r)] + V(|\rho(r)|) \Psi(r) = E \Psi(r).
\]

The total energy-functional having the contribution of this effective potentials is given by

\[
E = -\sum_{\lambda \in \text{occ}} \int dr \nabla \phi_{\lambda}(r) \nabla \lambda \phi_{\lambda}(r) + \int dr \rho(r)V_{xc}(r)
+ \frac{1}{2} \int dr \int dr' \rho(r)\rho(r') \left( \frac{\rho(r) - \rho(r')}{|r - r'|} \right) + E_{xc}[\rho(r)]
\]

where the 1st term is kinetic energy. The 2nd, 3rd, and 4th terms involve the external ion-electron \([ie]\), the Hartree, and exchange-correlation \([XC]\) contributions, respectively, which combine to provide the self-consistently determined effective potential, \( V_{\text{eff}}(r) \), written as

\[
V_{\text{eff}}(r) = V_H[\rho(r)] + V_{\text{xc}}[\rho(r)] + V_{\text{xc}}[\rho(r)].
\]

Generally, \( V_{\text{xc}}[\rho(r)] \) is given by the variation of the \( E_{\text{xc}}[\rho(r)] \) with respect to density \( \rho(r) \). The LDA, first introduced by Slater, is based on homogeneous electron gas.\(^{53}\) It simplifies the non-local exchange energy in terms of local \( \rho^{1/3} \) potential,\(^{54}\) as \( E_{\text{xc}}^{\text{LDA}} \simeq \int \rho(r) \epsilon_{\text{xc}}[\rho(r)] dr \). In semi-local form, it is simplified as \( E_{\text{xc}}^{\text{semi}} \simeq \int \rho(r) \epsilon_{\text{xc}}[\rho(r); \nabla^{(n)} \rho(r)] dr \) in terms of local XC energy density. While calculations using DFT potentials from LDA/GGA are the most successful and widespread in the last three decades, most of these calculations have resulted in significant underestimation of the bandgaps of semiconductors and insulators.

Many well-established DFT approximations fail to reproduce the observed energy gaps of semiconducting material due to wrong asymptotic behavior at \( r \to 0 \) and \( r \to \infty \). The LDA potential decays exponentially similar to density for finite systems, however, the potential should decay as \( -1/r \), i.e., Coulomb-like.\(^{55}\) Becke proposed a nonlocal correction to the energy-functional and then found the exchange potential,\(^{27}\) which gave a correct asymptotic form of exchange-energy density but failed to reproduce the exact behavior for the potential. Perdew and Wang developed GGA functionals.\(^{9,56}\) Later, van Leeuwen and Baerends provides a similar type of correction to LDA exchange, significantly improving the eigenvalues for the highest-occupied orbitals of atoms.\(^{25,57-59}\)

In short, the vLB-corrected LDA potential can be written as

\[
v_{\text{xc}}(r) = \left[ v_{\text{xc}}^{\text{LDA}}(r) + v_{\text{xc}}^{\text{vLB}}(r) \right] + v_{\text{xc}}^{\text{LDA}}(r),
\]

where, using \( z = \frac{\nabla \rho(r)}{\rho^{1/3}(r)} \) and \( \beta = 0.05 \),

\[
v_{\text{xc}}^{\text{vLB}}(r) = -\beta \rho^{1/3}(r) \frac{z^2}{1 + 3\beta z \sinh^{-1}(z)}
\]

This method of gradient correction to the XC-functional is more built into the theory and less artificial than self-interaction-corrected (SIC) approaches.\(^{35,37}\)

Singh, et al.\(^{35,37}\) implemented and discussed results for LDA+\( v \)LB within LMTO-ASA, in particular, the importance of the vLB-matching condition in the interstitial (the “asymptotic” value within a solid). That is, different electronic-structure (e.g., site-centered basis vs. plane-wave basis) methods have different reference zeroes for the potentials, i.e., \( v_0 \). Notably, here, any site-centered method can choose its potential zero inside the crystal, typically in the interstitial to define the stationary wave. As such, we set the LDA+\( v \)LB potential to \( v_0 \) and enforce its “asymptotic” \(( -1/r ) \) behavior in the interstitial. Hence, all potentials (i.e., Eq. 6) used to solve the KS equations and eigenvalues (dispersion) are defined relative to \( v_0 \). And, notably, we only require differences of \( I \) and \( A \) in reference to \( v_0 \), which cancels out. For approximate densities/potentials, a variational form of \( v_0 \) is useful, as it approaches (to second-order in density error) the KS kinetic energy (dispersion) of exact full-potential results.\(^{52}\) As noted in Ref. 35, and as discussed in our Results Section, changing the reference to atomic zero, say, to compare to plane-wave results, one requires the work (dielectric) function \( W \propto \epsilon \). Hence, an advantage of the site-centered basis is that it avoids the calculation of \( \epsilon \) for some quantities like bandgap, but recovers the same result. In any case, LDA+\( v \)LB can be implemented using, e.g., LibXC software,\(^{61}\) if the asymptotic behavior of \( V_{\text{eff}}(r \to s) \) is set judiciously in the interstitial of the solid.
FIG. 2. (Color online) vLB-exchange $v^{LB}_{\text{work}}$. The vLB correction is motivated by Becke, in terms of the shape of the potentials considered in this work. The vLB correction is motivated by Becke, which also leads to correct $-1/r$ behavior 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. This shows that an improved XC potential that is particularly accurate in the outer regions of the AS gives significantly improved results over the LDA.

The correctness of the eigenvalue of the highest occupied state, the appropriate asymptotic behavior and integer discontinuity of the exchange-correlation potential are all inter-related, which helps to estimate the accurate bandgaps in DFT methods. It is well known that Harbola-Sahani exchange-only functional gives accurate uppermost eigenvalues and correct asymptotic behavior for the finite systems too. Similar to the inbuilt derivative discontinuity of exact-exchange and Harbola-Sahani exchange-only potential, the vLB also shows similar discontinuities at the “shell-steps” in potential (Fig 2). Thus, it is not a coincidence that for a large number of systems the exact-exchange, self-interaction correction, and Harbola-Sahani exchange-only potential give significantly improved bandgaps compared to semi-local functionals. For Ge and Si, we plot (Fig. 2) the potential in the atomic sphere. The result provides some guidance when choosing (within the KS method) an exchange-correlation potential that is adequate for the problem at hand, e.g., bandgap calculation in our case. Trends in the results could be understood by comparing the shape of the potentials.

The discontinuous shift in the exchange potential turns out to be closely related to the “step structure” of the exchange potential of atoms, also shown by van Leeuwen-Baerends in the proposed model potential. Such a step is present regardless of how small the occupancy of a shell is, as long as it is greater than zero. Hence, in the fractionally-filled orbital case, the orbitals are filled with a successively increasing fractional particle number and a new step is created at the exact point when a new shell is opened. If a boundary condition $v_X \to 0$ as $r \to \infty$ is enforced for the potential, it shifts the whole potential discontinuously. Krieger, Li, and Iafrate observed that with just a tiny fraction of occupancy in a new shell, the exchange potential is shifted discontinuously, where an exact-exchange potential (e.g., van Leeuwen-Baerends) shows the required potential discontinuity. Similar considerations also apply for the time-dependent extension of DFT.

IV. RESULTS AND DISCUSSION

We demonstrate that improving the LDA for its behavior in the outer regions of the atomic sphere and basis set indeed leads to significant improvement in bandgap. We exemplify this by calculating bandgaps of Si, Ge, GaP, GaSb, GaAs, InAs, InSb and InP and compared them with existing theory and experiments. For better understanding, we discuss Ge and GaAs results in detail.

Bulk Germanium: In Fig. 3, we plot the band structure and total density of states (DOS) for Ge. Ge crystalizes in the diamond ($F\bar{d}3m$) structure. Two equivalent Ge’s are placed at $(0,0,0)$ and $(1/4,1/4,1/4)$ in a cubic cell with empty spheres (ES) at $(-1/4, -1/4, -1/4)$ and $(1/2, 1/2, 1/2)$. Both TB-LMTO-ASA and FP-NMTO use average Wigner-Seitz radius $1.387 \, \text{Å}$ for Ge. For FP-NMTO, the hard sphere radius is $0.971 \, \text{Å}$. The FP-NMTO-LDA gives a direct bandgap of $0.19 \, \text{eV}$ at $\Gamma$, which shows small improvement over LMTO+LDA of $0.10 \, \text{eV}$. With the interstitial region treated in the same way, use of the improved FP-NMTO basis set gets better values. However, if we switch to FP-NMTO-PBE, the bandgap is improved to $0.33 \, \text{eV}$, but the nature of bandgap remains direct, and more than $50\%$ underestimated with respect to the experiments ($0.74 \, \text{eV}$). In contrast, the FP-NMTO using LDA+vLB increases the bandgap to $0.86 \, \text{eV}$. The key point is that the correct nature of bandgap, which is indirect along $\Gamma \rightarrow L$, now matches experiment. The valence bands remain almost same, whereas the conduction bands shift away from $E_F$.

For Ge in LDA and GGA, see Fig. 3, shows no key difference in band structure. The spaghetti of bands originated from the Ge-4$s$ level were seen between $-13 \, \text{eV}$ and $-8.7 \, \text{eV}$. The Ge-4$p$ levels with a contribution of the Ge-4$s$ level toward the less binding energies are found at the region closer to the Fermi energy ($E_F$). These bands extend from $-8.7 \, \text{eV}$ up to $0 \, \text{eV}$. The conduction band mainly results from Ge-4$p$ with minor contributions of Ge-4$s$ atomic orbitals toward the least excited states of Ge-4$d$ from approximately $6.30 \, \text{eV}$ and up. Some degree
of degeneracy is also observed depending on which direction of the BZ is traversed: in the $L \rightarrow \Gamma$, and $\Gamma \rightarrow X$ direction. These degenerate bands belong to Ge and filled with Ge-$p$ electrons. The $\Gamma$ point has three degenerate bands; and the $X$ point at the edge of the BZ shows two values of energies with a twofold degeneracy.

Our results show that the fundamental gap of Ge is an indirect one with the maximum of the valence band occurring at $\Gamma$ and the minimum of the conduction band at the $L$ point. The calculated indirect gap at equilibrium lattice constant is 0.86 eV (versus 0.74 eV in experiments). The calculated direct gap at the $\Gamma$ point is $\sim 2.0$ eV. The LDA and GGA lead to the similar direct gaps of 0.19 eV and 0.33 eV, respectively. The LDA+vLB structural properties and bandgaps are in better agreement with experiments. This also shows improvement over beyond-DFT approaches.

As the vLB potential was not derived as a functional derivative, the corresponding energy functional is not yet known, as is so with other functionals. However, the exchange energy can be evaluated by applying the virial theorem based Levy-Perdew sum rule, and with this the prescription for performing DFT calculations for structural minimization for a vLB potential is complete.

The FP-NMTO calculated DOS, compared in Fig. 3 (right-panel), for vLB-corrected shows differences both in the energy positions and in the valence bandwidths with respect to LDA and PBE. We can see a weak shoulder at $-0.60$ eV followed by relatively broader peaks at $-1.35$, $-2.25$, and $-2.9$ eV and a sharper peak at $-4.7$ eV. In the conduction bands, a small shoulder can be found at 0.9 eV followed by a peak at 2.50 eV, with other peaks at 3.2, 4.0, and 4.8 eV. These DOS peak positions are in closer agreement with experiments. Although Ge is known to be an $sp$ material, a correct treatment of 3$d$ states is critical for accurate results. We found that major contribution of Ge-$3d$ states goes to conduction bands with smaller contribution to valence states. We consider Ge-$1s$+$2s$+$2p$ into the core, which leads to a better description of the band energies. Clearly, improving the basis set by including Ge-3$d$ states into valence achieves better energy description and (indirect) gap of Ge.

To further elaborate why vLB-corrected LDA in FP-NMTO produces better result, we plot XC-potential surface for LDA and LDA+vLB of Ge [100] plane in bottom panel of Fig 3. The $z$-axis represents the strength of XC-potential in the range $\{0, -3\}$. As the nature of interstitial region can not be understood for the full range of XC-potential, we show a truncated plot. At atomic sites, the XC-potential forms deep wells, and falls exponentially to 0 in LDA (Fig.3(a)). However, the LDA+vLB shows much slower fall than LDA (Fig.3(c)). The curvature of the XC-potential surface also improves accordingly. The plots are done at isosurface values of $-0.45$ eV/C for both LDA and LDA+vLB in Fig.3(b),(d). This shows that the XC-potential for Ge in LDA is much more localized at the atoms, but LDA+vLB is more spread over the cell.

Bulk Gallium-arsenide: GaAs crystallizes in the zinc-
<table>
<thead>
<tr>
<th>System</th>
<th>Si</th>
<th>Ge</th>
<th>GaP</th>
<th>GaSb</th>
<th>GaAs</th>
<th>InAs</th>
<th>InSb</th>
<th>InP</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(Å) (→)</td>
<td>5.50</td>
<td>5.57</td>
<td>5.45</td>
<td>6.00</td>
<td>5.65</td>
<td>6.04</td>
<td>6.48</td>
<td>5.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>XC(↓)</th>
<th>Band Gap (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA (LMTO)</td>
<td>0.49(I) 0.10 1.67 0.52 0.08 0.01 0.06 0.6</td>
</tr>
<tr>
<td>vLB (LMTO)</td>
<td>1.21(I) 0.06 1.46 0.05 0.04 0.01 0.01 0.5</td>
</tr>
<tr>
<td>LDA (NMTO)</td>
<td>0.79(I) 0.19(D) 1.53 0.25 0.33 0.01 0.00 0.5</td>
</tr>
<tr>
<td>PBE (NMTO)</td>
<td>0.85(I) 0.33(D) 1.68 0.51 0.54 0.04 0.02 0.90</td>
</tr>
<tr>
<td>vLB (NMTO)</td>
<td>1.25(I) 0.86(I) 1.87 0.94 1.43 0.39 0.79 1.18</td>
</tr>
<tr>
<td>Experiment</td>
<td>1.17(I) 0.74(I) 2.32 0.81 1.52 0.43 0.23 1.42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ES</th>
<th>Band Gap (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mBJ28–32</td>
<td>1.15 0.83 2.25 0.95 1.53-1.64 0.42-0.67 0.25-0.47 1.42-1.62</td>
</tr>
<tr>
<td>LB9484</td>
<td>0.25 0.00 0.61 0.00 0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td>GW72,82–84</td>
<td>1.29-1.31 0.65-0.71 2.80 0.62 1.58 0.31 0.08 1.44</td>
</tr>
<tr>
<td>HGH85</td>
<td>– – 2.44 0.55 1.01 0.19 0.21 1.23</td>
</tr>
<tr>
<td>GDF86</td>
<td>1.17 0.49 2.47 0.93 1.72 1.40 0.99 2.55</td>
</tr>
<tr>
<td>HSE87,88</td>
<td>1.28 0.00 2.47 0.72 1.21 0.39 0.29 1.64</td>
</tr>
<tr>
<td>TPSS89</td>
<td>0.82 0.56 1.98 0.08 0.52 0.00 0.00 0.90</td>
</tr>
<tr>
<td>EXX91</td>
<td>1.50 1.01 – – 1.82 – – –</td>
</tr>
</tbody>
</table>

TABLE I. FP-NMTO bandgaps for group IV and III-V materials (in eV) from LDA, PBE and LDA+vLB compared with other theories and experiments. For LMTO-vLB see ref. [35]. I=indirect-bandgap; D=direct-bandgap.

blende (F43m) structure, where Ga and As-atoms occupy (0,0,0) and (1/4,1/4,1/4) sites, respectively. The ES are placed to maintain the curvature of potential. We plot FP-NMTO calculated band structure and DOS in Fig. 4(a). The bandgap from LDA, PBE and LDA+vLB comes out to be 0.33 eV, 0.54 eV, and 1.43 eV, respectively. The FP-LMTO+vLB calculated bandgap of 1.43 eV shows good agreement with experimentally observed bandgap of 1.52 eV. Clearly, LDA and PBE hugely underestimate the bandgap compared to LDA+vLB. Moreover, the direct (D) nature of the bandgap is also well reproduced. The band structure and DOS in Fig. 4 does not show any constant shift, which rules out any scissor operator behavior of LDA+vLB. The band structure shows larger improvement in energy levels at Γ than at X or L points. We see no change in shape of DOS except at the conduction bands near $E_F$ and valence bands at $-2.25$ eV. In Fig. 4 (b),(c), the potential-energy surface for GaAs is shown along [100], where overlap between the potential spheres are determined self consistently. The $z$-axis shows the strength of XC potential. For clarity, we show a truncated XC-potential surface in the range (0,−3) along z-axis. We can see that at atomic sites, the XC potential forms deep well like structure, and falls exponentially to 0, which is similar to Ge. The light yellow zone in Fig. 4(c), shows smaller interstitial contribution. This is only possible if we have smaller number of interstitial electrons than LDA (darker spots), which consequently helps correcting the energy levels in conduction band.

We summarize the bandgaps of IV and III-V semiconductors in Table I. The results show that improving the basis set (localized, compact and complete) in conjunction with LDA+vLB yields significantly better results, especially in comparison to modified Becke-Johnson (mBJ), GW, and hybrid functionals. So, a better localized basis with no approximations (Linear vs N-th-order MTO) is important, but the improve description of exchange and it asymptotic behavior is equally necessary (LDA vs LDA+vLB).

Comparing the various results in Table I, it is evident that the improvement in basis set affects significantly the results – for example, compare Si, Ge, and GaSb between LDA and LDA+vLB and LMTO-ASA vs FP-NMTO. (Keep in mind that LMTO-ASA has always used ES to improve the spherical basis for semiconductor materials, which is why it does alright.) In all cases, the improved basis set is critical; although, when combined with vLB correction in the interstitial, the results approach those of optimized mBJ, as well as GW and hybrid functionals. For example, vLB does not improve the gap for Ge within LMTO-ASA but the FP-NMTO basis set dramatically improves the gap with LDA+vLB, and gets correctly an indirect gap.

For completeness, we note that we find a large different between results from LDA+vLB (present calculation) and related LB94 (Ref. 29), which used a linear-augmented plane-wave code (Wien2K) with local atomic orbitals. As similar results are found from vLB and mBJ (which also uses Wien2K), we believe that this difference arises from the implementation of vLB asymptotic conditions and our different basis sets, as discussed in section III and section IV B regarding plane-wave results.
A. Application to 2D materials: h−BN and h−SiC

The 2D materials, like of graphene and hexagonal boron nitride (h−BN), have drawn tremendous attention in terms of both fundamental physics and possible applications in energy-generation devices. Single layers of graphene and h−BN have been fabricated and found to be stable at room temperature. Here, we exemplify the efficacy of our approach by discussing the electronic structure of 2D h−BN and SiC.

2D h−BN is a group III-IV binary compound displaying interesting chemical and electronic structure properties that leads to wide range of technological applications, such as protective coating, deep ultraviolet emitter, and transparent membrane. 2D h−BN has similar honeycomb atomic structure as graphene with a lattice mismatch of 2%, where each primitive cell contains a B and N atom and share total of eight electrons. Each B (N) atom forms sp²-hybridized state with N (B), where each state forms three in-plane σ (6-electrons) and an out-of-plane π (2-electrons) bond. The NMTO-vLB energy minimized lattice constant of 2.508 Å is in good agreement with that measured (2.504 Å). Unlike graphene, h−BN is a wide bandgap insulator. The calculated bandgap of ~4.5 eV, in Fig. 5(b), falls within the experimentally observed bandgap range of 3.6−6 eV. The nature of bandgap can not be clearly stated as the band-energies of the lowest-occupied bands at K and M are the same, i.e., a flat band along (K-M).

Similarly, 2D h−SiC is a group IV binary compound that can be viewed as graphene (2D C) or silicene (2D Si) doped with ‘Si’ or ‘C’, respectively. The honeycomb lattice of randomly but homogeneously distributed Si atoms in symmetric, semi-metallic graphene opens up a large gap. Because of its wide bandgap, h−SiC band structure has been in active study for optoelectronic applications. Unlike the polymorphs of carbon, h−SiC is a polar material. In spite of the fact that both constituents of SiC are Group IV elements, charge is transferred from Si to C due its to higher electronegativity.

The NMTO-vLB energy minimized lattice constant of 2.62 Å is in good agreement with the measured a=2.6±0.2 Å. The bandgap of 3.90 eV (indirect along K to M in Fig. 5(c)) is underestimated compared to experiment (2.96 eV), a trend is observed in most other theories. We also found good agreement in structural properties with experiments, while other theoretical results show significant differences, which yield a gap to be as high as 4.42 eV due to strong excitonic effects included at the G0W0 level of theory.

B. Comparing vLB-corrected Results

To avoid confusion when comparing results between different (e.g., site-centered basis vs. plane-wave basis) methods and various implementation of modified semi-local functionals (like vLB) or optimized, screened-range-separated hybrid functionals, recall that each method has different potential zeroes, v₀. Plane-wave methods use a global reference set to atomic zero (v₀ = 0) far outside the atom or crystal, which requires a dielectric function, ε, to solve the macroscopic electrostatics, in particular, to set the asymptotic (−1/r) condition in hybrid functional, two range-partition variables must obey a sum rule (α + β)r−1 = ε−1r−1, where ε = 1 for an atom in vacuum and 1 ≤ ε ≤ ∞ for a crystal. Optimal tuning depends on the values of α, β, and ε. Hence, to get the excited states or bandgap, you must calculate ε, a significantly more costly calculation. In contrast, all-electron, site-centered methods can choose to take the potential zero inside the crystal interstitial (eigenvalues relative to v₀) and set the potentials asymptotic behavior there. Then only require I − A, which is independent of v₀. To change the reference to atomic zero to compare, say, to plane-wave results, one requires the work function W ∝ ε. Hence, site-centered basis can avoid the calculation of ε for bandgaps still obtain the same result as found from plane-wave calculations. For implementations using other methods, like linear-augmented-plane-waves with local atomic orbitals, it is judicious to de-
fine the asymptotic value of the potential within the solid. How the asymptotic behavior is addressed affects results, as found, for example, for LB94, LDA, and LDA+vLB. Notably, LDA+vLB should approach the mBJ, GW, and screened-hybrid functional results, as we indeed found in Table I. Of course, care is needed in comparing results in the solid-state limit.

V. CONCLUSION

The bandgaps of IV and III-V semiconductors are underestimated by most semi-local exchange-correlation functionals. To address the problem, we combine (a) self-consistent FP-NMTO method [providing a compact and complete optimal basis throughout a supercell volume – with no approximations to interstitial regions] with (b) vLB correction to the LDA exchange. The FP-LMTO-vLB approach provides a fast and efficient way to calculate accurate bandgaps, comparable to hybrid exchange functionals, with the speed of semilocal functionals (e.g., GGA). The combined improvement in optimal basis set (FP-NMTO) and asymptotic behavior of the exchange hole (LDA+vLB) yield much better bandgaps. The self-consistent FP-NMTO-vLB can be more effective in calculating material properties (e.g., electronic, optical), localized atomic orbitals (Wannier functions), or real-space tight-binding parameters for electronic-structure studies, as exemplified here for bandgaps in group IV and III-V semiconductors and 2D materials.

VI. ACKNOWLEDGEMENTS

S.D. and A.M. would like to thank Yoshiro Nohara and O. K. Andersen for permission to modify the FP-NMTO. S.D. is thankful to I. Dasgupta for useful discussion. Research at Ames Laboratory was funded by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences, Materials Science and Engineering Division, which is operated for the U.S. DOE by Iowa State University under contract DE-AC02-07CH11358.

APPENDIX:

Reproducing exact-exchange: The H atom

The vLB potential goes as $-1/r$ for densities that decay exponentially as a function of $r$, i.e., as $\exp(-r)$, for positions far from the origin. The effective potential $V_x + V_c + V_{LB}$ along with the LDA potential $V_x + V_c$ are shown in Fig. 6 for an electron in a hydrogen atom. The exact known densities for these systems are used to obtain these potentials. The potentials are compared with the exact XC potential given by negative of the electrostatic potential for these single-electron systems. Except close to $r = 0$, it is clear that the vLB-corrected potential is very close to the exact exchange-correlation potential. In Table III, we display the exact potential (self-interaction free) for a single-electron in a hydrogen (H) atom and compare exact energies calculated with exact, LDA and LB-corrected LDA potentials $V_{exact}$ potentials, where a much smaller error in self-interaction energy of the vLB-corrected potential is notable. Clearly, the LB-corrected LDA potential provides a significant improvement to the energy without invoking need for self-interaction corrections.

FIG. 6. (Color online). LDA+vLB compared with the exact-exchange and LDA potentials for a single-electron density. LDA+vLB matches closely the exact exchange-correlation.

<table>
<thead>
<tr>
<th>$V_{exact}$</th>
<th>$V_{LB}$</th>
<th>$V_{LDA}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>0.3144</td>
<td>0.2858</td>
</tr>
<tr>
<td>%error</td>
<td>9.09</td>
<td>29.20</td>
</tr>
</tbody>
</table>

TABLE II. Total energy for 1$e^-$ in a H atom calculated using exact, LDA and LDA+vLB potentials. LDA+vLB yields significant improvement to energy without invoking self-interaction corrections. Error is given as $\frac{V_{exact} - V}{V_{exact}} \times 100\%$.

Calculation Details –TB-LMTO-ASA

In TB-LMTO-ASA core states are treated as atomic-like in a frozen-core approximation and energetically higher-lying valence states are addressed in the self-consistent calculations of the effective crystal potential, which is constructed by overlapping Wigner-Seitz spheres for each atom in the unit cell. Two-fold criteria for generating crystal potential, on the same footings of TB-LMTO-ASA, has been used: (a) use of trial wave function, i.e., linear combinations of basis functions like plane waves in the nearly free-electron method, and (b) use of matching condition for partial waves at the muffin-tin sphere. We have used the LDA correlation parameterized by van Barth and Hedin with corrected vLB exchange, matched at the ASA radii. Following atomic-sphere-approximation in TB-LMTO-ASA, the open
shell structured semiconductors are filled with empty spheres (ESs) for improved basis. Here, ESs are empty sites with no cores and small density of electronic charges. The dependence of dimensionless parameter $x$ present in vLB correction on $R_{ASA}$ is very crucial, so appropriate choice of $R_{ASA}$ is important. In all calculations, we chose $R_{ASA}$ by $\pm 5\text{-}10\%$ from default values to control the overlapping of atomic spheres and empty spheres to reduce the loss of electrons into the (unrepresented) interstitial for open-shell structures, e.g. semiconductors.

Timings Comparison for FP-NMTO (vLB) and Quantum Espresso (HSE06)

To given some comparison of times, we used FP-NMTO-vLB and Quantum Espresso HSE06 for two cases, Ge and 2D-BN, as shown in Table III. Even though the methods are quite different (i.e., basis sets, etc.) the hybrid functional is severely costly, in particular a second q-mesh is required for solution.

<table>
<thead>
<tr>
<th>System/Method</th>
<th>HSE06 (QE)</th>
<th>vLB (NMTO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ge (k-mesh 8x8x8)</td>
<td>250.8</td>
<td>6.60</td>
</tr>
<tr>
<td>2D-BN (k-mesh 12x12x4)</td>
<td>69.6</td>
<td>2.76</td>
</tr>
</tbody>
</table>

TABLE III. Time (in minutes) for two representative cases: Ge and 2D-BN. We use HSE06 in Quantum Espresso (QE) and LDA+vLB in NMTO on same machine for one-one comparison between two methods.

Wannier Representation in FP-NMTO

In Fig. 7, we show a Wannier function representation for the 2D-BN as found within the FP-NMTO using LDA-vLB.

Advantage of FP-NMTO over TB-LMTO-ASA

The improved basis set of FP-NMTO method is accurate, minimal and flexible. Accurate because the FP-NMTO basis solves the KS equation exactly for overlapping muffin-tin potentials. Orthonormalized FP-NMTOs are localized atom-centered Wannier functions, generated in real space with Green-function techniques, without projection from band states. The TB-LMTO-ASA lacks in all above. Increased computational time is the only downside of the FP-NMTO compared to TB-LMTO-ASA.

---