Better band gaps for wide-gap semiconductors from a locally corrected exchange-correlation potential that nearly eliminates self-interaction errors

Prashant Singh\textsuperscript{1}, Manoj K Harbola\textsuperscript{2} and Duane D Johnson\textsuperscript{1,3}

\textsuperscript{1} Ames Laboratory, U.S. Department of Energy, Iowa State University, Ames, Iowa 50011-3020, United States of America
\textsuperscript{2} Department of Physics, Indian Institute of Technology, Kanpur, 208016, India
\textsuperscript{3} Materials Science & Engineering, Iowa State University, Ames, Iowa 50011-2300, United States of America

E-mail: prashant@ameslab.gov, mkh@iitk.ac.in, ddj@iastate.edu and ddj@ameslab.gov

Received 25 April 2017, revised 31 July 2017
Accepted for publication 2 August 2017
Published 8 September 2017

Abstract
This work constitutes a comprehensive and improved account of electronic-structure and mechanical properties of silicon-nitride (\(\text{Si}_3\text{N}_4\)) polymorphs via van Leeuwen and Baerends (LB) exchange-corrected local density approximation (LDA) that enforces the exact exchange potential asymptotic behavior. The calculated lattice constant, bulk modulus, and electronic band structure of \(\text{Si}_3\text{N}_4\) polymorphs are in good agreement with experimental results. We also show that, for a single electron in a hydrogen atom, spherical well, or harmonic oscillator, the LB-corrected LDA reduces the (self-interaction) error to exact total energy to \(\sim 10\%\), a factor of three to four lower than standard LDA, due to a dramatically improved representation of the exchange-potential.

Keywords: wide-band semiconductors, Van-Leeuwen Barends correction, self-interaction correction

(Some figures may appear in colour only in the online journal)

1. Introduction
The wide-band gap silicon nitride (\(\text{Si}_3\text{N}_4\)) is an important engineering ceramic material owing to its superior physical and chemical properties, high chemical and thermal stability, and good wear resistance. It has been widely applied in cutting tools, engine components, ball bearings and gas turbines [1–4]. The \(\alpha\)-\(\text{Si}_3\text{N}_4\) and \(\beta\)-\(\text{Si}_3\text{N}_4\) can be obtained using conventional synthesis conditions. Both phases are in the hexagonal crystal system. In 1999, the cubic spinel structure (\(\gamma\)-\(\text{Si}_3\text{N}_4\)) was synthesized by applying high pressure with high temperature [4]. The superior physical and chemical properties of wide-gap semiconductors make these materials an ideal choice for device fabrication for many applications, including light emitters, high-temperature and high-power electronics and microwave devices, and micro-electromechanical systems. The effects of temperature on these materials and devices still remain of critical interest [5, 6]. In addition, wide-gap materials are being considered as alternative gate oxides in the electronic industry, and for low-loss dielectric waveguides in integrated optics [4], which also find its use in light-emitting devices too [7–9].

Here, we present an electronic-structure study of wide-band-gap silicon-nitride polymorphs. We use spin-polarized version of the van Leeuwen and Baerends (LB) [10] exchange correction to local density approximation (LDA) [11]. The LB-corrected LDA is implemented within the framework of tight-binding linear muffin-tin orbital method [12, 13]. The LB correction [10] to the exchange potential corrects the asymptotic regions of finite systems, such as the potential
The crystal structure and mechanical properties of... corrections [17] improved treatment without explicit need for self-interaction [13]. We discuss how the better exchange-correlation provides large improvement over LDA and GGA calculations, and of silicon-nitrides, the LB-corrected LDA shows significantly-Sahni potential [16]. For polymorphs of calculations employing the optimized potential method... between atoms [13]. Expectedly, this improves the valence... like plane waves in the nearly free-electron method, and... consisting of layers of Si and nitrogen atoms in the sequence behaves as $-1/r$ at large $r$. For an extended system, asymptotic behavior is expected within the interstitial region between atoms [13]. Expectedly, this improves the valence and conduction band energies of an extended system leading to better band gaps [13, 14]. The idea is based on the results of calculations employing the optimized potential method [15] and the Harbola–Sahni potential [16]. For polymorphs of silicon-nitrides, the LB-corrected LDA shows significantly large improvement over LDA and GGA calculations, and found to be similar to hybrid and exact exchange functions [13]. We discuss how the better exchange-correlation provides improved treatment without explicit need for self-interaction corrections [17–19].

In section 2, we describe computational methodology. The crystal structure and mechanical properties of Si$_3$N$_4$ are discussed in section 3. The electronic-structure properties of are discussed in section 4. We summarize our results in section 5, and provide an appendix with a discussion of errors and improvements in self-interaction in the LDA and the LB-corrected LDA for three single-electron examples.

2. Computational details

Density functional theory calculations are performed using the TB-LMTO method within the atomic sphere approximation (ASA) [12]. In TB-LMTO, core states are treated as atomic-like in a frozen-core approximation and higher-lying valence states are addressed in the self-consistent-field method applied to the effective crystal potential. This is constructed by overlapping spheres for each atom in the unit cell. The TB-LMTO uses two fold criteria for generating crystal potential: (a) a linear combination of basis functions like plane waves in the nearly free-electron method, and (b) matching condition for partial waves at the muffin-tin sphere [12]. The open-shell structure of semiconductors and the use of the ASA within the TB-LMTO [12] requires us to divide unit cell into atomic spheres (AS) and empty spheres (ES) to fill space for improved basis sets. Here, ES are sites with no cores and small density of electronic charges.

Next, we utilize the LB-correction to LDA-exchange potential [10] to improve the local potential associated with the $-1/r$ behavior of the exchange potential at large distances from the atomic core, which is the interstitial region between neighboring atoms. The LB-corrected LDA provides a dramatic improvement in the band gaps, as well as how it compares, and its relation to, pseudopotential methods that require, in addition, the static dielectric constant [13, 14]. Details and further background are available [16, 20, 21]. The contribution of exchange in empty spheres is very small; thus, any correction to exchange will be even smaller, so we implement the LB-correction in the ASA spheres only. Details of the LDA potentials (correlation $V_c$ and exchange $V_x$) and the LB-correction to these are given in the appendix for a single-electron system for potentials with different boundary conditions. There we identify and clarify the origin of the improvements to total energy of the system, reducing the error in energy from LB-corrected LDA to less than 10% compared to 30% for LDA.

All calculations presented here are done self-consistently (utilizing Anderson mixing [22] of charge densities) and non-relativistically for given experimental geometry until the ‘averaged relative error’ to the previous iteration reaches $10^{-5}$ for the charge density and $10^{-4}$ Ryd/atom for the energy. We use LDA+LB corrected exchange in combination with correlation potential parametrized by van Barth-Hedin [11]. The k-space integration over occupied electronic states is performed using the tetrahedron method on a $12 \times 12 \times 12$ for cubic cells, $6 \times 6 \times 12$, and $6 \times 6 \times 18$ set of k-point mesh. In all calculations we chose $R_{ASA}$ to be within 5% to 10% of the default value [12].

3. Crystal structure

We show three crystallographic structures of Si$_3$N$_4$ in figure 1, designated as $\alpha$, $\beta$ and $\gamma$-phases [23]. The $\alpha$ and $\beta$ phases are the most common forms, which can be produced under normal pressure condition. The $\gamma$ phase can only be synthesized under high pressures and temperatures and has a hardness of 35 GPa [23, 24]. The crystal structures and unit-cell parameters of $\alpha$, $\beta$, and $\gamma$-Si$_3$N$_4$ have been extensively studied by many [25–35].

The $\alpha$- and $\beta$-phase have trigonal (Pearson symbol $hP28$, space group P31c, No. 159), and hexagonal ($hP14$, $P6_3$, No. 173) structures, respectively, which are built up by corner-sharing SiN$_4$ tetrahedra. They can be regarded as consisting of layers of Si and nitrogen atoms in the sequence
Table 1. Lattice constant and Bulk modulus of Si₃N₄ polymorphs obtained from a Birch–Murnaghan equation of state analysis of data points calculated using LB-corrected LDA exchange-correlation potential. We compare LB calculated result with experiments [42–44] and other theories [42, 47, 48]. The calculated c/a of 0.73206 (α) and 0.36948 (β) compares well with experiments 0.72497 (α) and 0.38262 (β), respectively.

| Phase | α (Å) | β (Å) | γ (Å) | Expt. LB | LDA | Expt. LB | LDA |
|-------|-------|-------|-------|---------|-----|---------|-----|
| Si₃N₄ |       |       |       |         |     |         |     |
| α     | 7.752 | 7.819 | 7.586 | 248     | 275.6 | 257     |     |
| β     | 7.608 | 7.968 | 7.586 | 256     | 235.5 | 282     |     |
| γ     | 7.742 | 7.681 | 7.311 | 300     | 272.8 | 308     |     |

ABCDABCD… or ABAB… in α- and β-phase, respectively. The AB layer is the same in the α and β phases, and the CD layer in the α phase is related to AB by a c-glide plane. The tetrahedra’s in β-phase are interconnected in such a way that looks like tunnels, running parallel with the c-axis of the unit cell. Due to the c-glide plane that relates AB to CD, the α structure contains cavities instead of tunnels. An essential difference between the two phases is that the lattice constant in the c-axis of the lattice in α-phase is approximately twice that in β-phase [36, 37].

In figures 1(a) and (b), the Si–atom in α and β-phase coordinates with four N atoms, forming a Si₄ tetrahedron, however, a N–atom is coordinated with three Si-atoms, forming a NSi₃ triangle. These features suggest Si–sp³ and N–sp² hybridizations for the nearest silicon and nitrogen pairs [38].

The β-phase has spinel structure (β-phase space group Fd3m, No. 227), which is often designated as c-modification in boron nitride (c-BN) [4]. It has a spinel-type structure in which two Si-atoms each coordinate six nitrogen atoms octahedrally, and one Si atom coordinates four nitrogen atoms tetrahedrally [24]. The γ-phase is built up from a unit cell with 44 atoms located along Γ–M, Γ–A, and K–M lines in α–, β–, and γ–Si₃N₄, respectively, while bottom of conduction band is located along γ, at Γ, and along K–M, respectively. α- and β-phase show indirect band gap. Table 2 summarizes the calculated band gaps. For α and β-Si₃N₄, we find reasonably good agreement with the previous calculations [53]. This shows that better treatment of the exchange in the asymptotic region improves the band-gap significantly. For better description of the potential, we provide a discussion of the LB-correction in the appendix for simple single-electron models, which reduces the error from the exact total energy due to associated self-interaction.

In α–Si₃N₄, the direct band gap of 5.2 eV at Γ is comparable to that in β–Si₃N₄, which closely follow direct band-gap of γ–Si₃N₄ (~6.0 eV). Amorphous phase of Si₃N₄ are large band-gap materials while in previous calculations γ-phase has been reported to be smaller band gap system than other polymorphs. However in present calculation, we found γ-phase has band-gap of similar order as the other two polymorphs, and in agreement with experiments [40].

Comparing total electronic density of states (DOS) of α–, β–, and γ-phase shows that α–Si₃N₄ has a larger band gap, and a steeper VB/CB edge compared to the other two phases. In α–Si₃N₄, the bands around the Fermi energy form a so-called flat/steep bands. The flatness of the bands is very

Table 2. Structural data of Si₃N₄–polymorphs

| Polymorph | α (Å) | β (Å) | γ (Å) | Expt. LB | LDA | Expt. LB | LDA |
|-----------|-------|-------|-------|---------|-----|---------|-----|
| α         | 7.752 | 7.819 | 7.586 | 248     | 275.6 | 257     |     |
| β         | 7.608 | 7.968 | 7.586 | 256     | 235.5 | 282     |     |
| γ         | 7.742 | 7.681 | 7.311 | 300     | 272.8 | 308     |     |

4.2. Electronic structure of Si₃N₄ polymorphs

We calculate energetics of Si₃N₄–polymorphs at lattice constants shown in table 1. In terms of energetics, the α– and β–phases have very similar energies at the minimum, the energy difference between these two naturally occurring phases is around 30 meV per formula unit. However, the γ–phase has higher ground state energy than the other two phases. Out of the three crystalline silicon nitride phases, the β-phase has the lowest ground state energy, and is unambiguously thermodynamically most stable phase [4].

In figure 2, we show the band-structure and the density of states calculated at minimum lattice-constant determined by us using the LB-corrected LDA as described in section 4.1. The calculated band structures are in good agreement with measurements [40, 49–51]. The valence band width has two split regions separated by 3–4 eV, the lower and upper band-widths of the valence band are 4–5 eV and 10–12 eV, respectively. The calculated valence and conduction band-widths are in both quantitative and qualitative agreement with the measured valence and conduction bands from the x-ray photoemission [49–51], soft-x-ray emission [52], and bremsstrahlung isochromat-spectroscopy [51]. The top of the valence band occurs along K–M, Γ–A, and K–M lines in α–, β–, and γ–Si₃N₄, respectively, while bottom of conduction band is located along γ, at Γ, and along K–M, respectively. α- and β-phase show indirect band gap. Table 2 summarizes the calculated band gaps. For α and β-Si₃N₄, we find reasonably good agreement with the previous calculations [53]. This shows that better treatment of the exchange in the asymptotic region improves the band-gap significantly. For better description of the potential, we provide a discussion of the LB-correction in the appendix for simple single-electron models, which reduces the error from the exact total energy due to associated self-interaction.

In α–Si₃N₄, the direct band gap of 5.2 eV at Γ is comparable to that in β–Si₃N₄, which closely follow direct band-gap of γ–Si₃N₄ (~6.0 eV). Amorphous phase of Si₃N₄ are large band-gap materials while in previous calculations γ-phase has been reported to be smaller band gap system than other polymorphs. However in present calculation, we found γ-phase has band-gap of similar order as the other two polymorphs, and in agreement with experiments [40].

Comparing total electronic density of states (DOS) of α–, β–, and γ-phase shows that α–Si₃N₄ has a larger band gap, and a steeper VB/CB edge compared to the other two phases. In α–Si₃N₄, the bands around the Fermi energy form a so-called flat/steep bands. The flatness of the bands is very
sensitive to the lattice patterns and orbital interactions in real atomic structures. The polymorphs have very different multiple structures in the DOS in both the VB and CB regions. In figure 2, the electronic structure of silicon-nitride polymorphs split to form the separate valence band (near -10 eV; not shown in Fig.), which reflects the presence of smaller asymmetric part of the potential in Si3N4. This allows stronger mixing between the low-lying N and C states.

To investigate the bonding nature of α and β-Si3N4 in more detail, we compared the charge calculated around the Si and N sites. Here, the charges were calculated from the total electronic charges enclosed in atomic-spheres around Si/N-sites. The radii of the spheres were chosen so as to touch both spheres at the density minimum position of the Si–N bond. The calculated atomic charges around the inequivalent Si (Si1, Si2), and N (N1, N2, N3, N4) sites are (10.08, 10.18), and (7.52, 7.99, 7.89, 7.71), respectively, in primitive unit cell of α-Si3N4. Ideally, α-phase should have total of 56 e−s, however, we find 51.37 e− s on (Si, N) atomic spheres and 4.63 e− s go into empty-spheres (ESs). Around ~8% of charge goes into ESs. The electrons on Si–AS (N–AS) is lower (higher) than that of the neutral Si-atom (N-atom). The DFT valence charge density of α, β, and γ-Si3N4 have indicated the charge transfer from the silicon atoms to nitrogen atoms. This is mainly ascribed to the difference in the electronegativity, as the nitrogen atom is more electronegative than the silicon atoms. These results give direct experimental evidence for charge transfer from the Si to N atoms in Si3N4-polymer [53]. In the similar charge analysis for β- and γ-phase, we find ~7% (AS = 19.65 e−; ES = 1.45 e−) and ~6.1% (AS = 32.88 e−; ES = 2.12 e−) of electronic-charge goes on ESs. Keeping in mind 5–10% overlap of atomic-spheres the amount of charge found in ESs is not unexpected.

It has been suggested before [55, 56] that p − d π interaction between Si and N stabilizes the 120° Si–N–Si in α– and β−phase Si3N4 compared to the smaller bond angles found in other trivalent group-V systems. However, we do not find any significant Si-d orbital contribution in the valence-band states. Here, N being a first-row element provides possible explanation for the planar triply-coordinated N-site. The first-row elements with localized 2p states behave differently from the elements lying just below them in the periodic table due to the formation of energetically more favorable sp2-hybridized bands in some first-row elemental solids. The insulator silicon nitride forms crystals with predominantly covalent character and due to the presence of nitrogen, an element from group-III, which leads to hybrid orbitals between Si and N differing considerably from the original atomic-orbitals.

### 5. Conclusion

We have shown that the use of LB corrected LDA-exchange for the wide band-gap silicon-nitride polymorphs improves the structural-properties (e.g. lattice constant and bulk-modulus) and electronic-properties (e.g. the band-gap) significantly. The LB-corrected LDA exchange provides improvement to the exchange-potential over the entire region except near (r → 0). The LB unerringly corrects the self-interaction in the LDA without the need of using SIC-LDA. We have shown that in existing local-density based exchange-correlation potentials, the LB-correction can be used as an alternative scheme.

### Acknowledgments

The research at Ames Laboratory was supported by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences, Materials Science and Engineering Division. Ames Laboratory is operated for the U.S. DOE by Iowa State University under Contract No. DE-AC02-07CH11358.
Appendix

In the appendix, we describe the LDA potential employed in our calculations along with the LB correction to it. We show significant reduction in the self-interaction error in the LDA potentials that is brought about by the LB correction. The spin-polarized exchange and correlation potentials in the LDA are given in terms of the spin density $\rho_\sigma$ by the expressions

$$V_x = -\left(\frac{6\rho_\sigma}{\pi}\right)^{1/3}$$

(A.1)

and

$$V_c = -0.0254 \times \ln \left(1 + \frac{75.59}{r_s}\right)$$

(A.2)

where $r_s$ is the Wigner–Seitz radius corresponding to $\rho_\sigma$, $\sigma$ is spin-component. The LB-corrected potential is given in terms of the density and its gradient as

$$V_{LB}^{\alpha\beta}(r) = -\beta \rho_\sigma^{3/3} \frac{x_\alpha^2}{1 + 3\beta x_\sigma \sinh^{-1}(x_\beta)}.$$  

(A.3)

where $\beta = 0.05$ was used in the original formulation [10].

The variable $x_\sigma = |\nabla \rho_\sigma(r)|/\rho_\sigma^{4/3}(r)$ signifies the change in mean electronic distance provided density is a slowly varying function in given region with strong dependence on gradient of local radius of the atomic sphere $R_\text{ASA}$.

The LB potential goes as $-1/r$ for densities that decay exponentially as a function of $r$. The effective potential $V_x + V_c + V_{LB}$ along with the LDA potential $V_x + V_c$ are shown in figure A1 for an electron in a hydrogen atom, in a spherical well of depth 5 a.u. and radius 1 a.u. and in a harmonic potential with $\omega = 1.0$ a.u. We note that while the density of the electron in the first two systems decays as $\exp(-r)$ for positions far from the origin, in a harmonic oscillator the density decays as $\exp(-r^2)$. The exact known densities for these systems are used to obtain these potentials. The potentials are compared with the exact exchange-correlation potential given by negative of the electrostatic potential for these single-electron systems. It is clear that except close to $r = 0$, the LB-corrected potential is very close to the exact exchange-correlation potential for both the systems.

The proximity of the LB-corrected potential to the exact one also leads to much smaller error in self-interaction energy. In the table A1, we have displayed the exact energy (self-interaction free) for a single electron systems given by

$$E_{\text{exact}} = \frac{1}{2} \int \text{d}r V_{\text{exact}}(r) \rho(r) \quad (A.4)$$

and compared these calculated with approximate exchange-correlation functional i.e. $V_{\text{LDA}}$ or $V_{\text{LB}}$, for the same potentials shown in figure A1. To calculate the self-interaction energy in the approximate potentials, we have used the expression in A.4 instead of the exchange-correlation energy functional. This is because for single electron systems the exact exchange-correlation energy is indeed given by A.4.

In table A1, the error due to the self-interaction in the LDA is about 30% in all three systems, compared to the LB-corrected results where it is reduced by a factor of three (to about 10%). The LB-corrected LDA shows significant improvement in total-energy without the need of separate treatment of self-interaction [17–19].

References

[1] Oba F, Tatsumi K, Tanaka I and Adachi H 2002 J. Am. Ceram. Soc. 85 97
[2] Oba F, Tatsumi K, Adachi H and Tanaka I 2001 Appl. Phys. Lett. 78 1577
[3] Zerr A, Kempf M, Schwarz M, Kroke E, Goken M and Riedel R 2002 J. Am. Ceram. Soc. 85 86
[4] Zerr A, Miehe G, Serghiou G, Schwarz M, Kroke E, Riedel R, Fueß H, Kroll P and Boehler R 1999 Nature 400 340
[5] Nuedeck P G, Okojie R S and Chien L-Y 2002 Proc. IEEE 90 1065
[6] Dreike P L, Fleetwood D M, King D B, Sprauer D C and Zipperian T E 1994 IEEE Trans. A 17 594
[7] Volodin V A, Efremov M D and Grisenko V A 1997 Solid State Phenom. 501 57
[8] Steimle R F, Sadd M, Muralidhar R, Rao R, Hradsky B, Straub S and White B E 2003 IEEE Trans. Nanotechnol. 2 335
[9] Choi S, Yang H, Chang M, Baek S, Hwang H, Jeon S, Kim J and Kim C 2005 Appl. Phys. Lett. 86 251901
[10] van Leeuwen R and Baerends E J 1994 Phys. Rev. A 49 2421
[11] van Barth U and Hedin L 1972 J. Phys. C: Solid State Phys. 15 1629

| System | $V_{\text{exact}}$ | $V_{\text{LB}}$ | % change | $V_{\text{LDA}}$ | % change |
|--------|------------------|----------------|----------|----------------|----------|
| H Atom | 0.3144           | 0.2858         | 9.09     | 0.2226         | 29.20    |
| Well   | 0.6240           | 0.5513         | 11.65    | 0.4052         | 35.06    |
| Oscillator | 0.4021 | 0.3720 | 7.49    | 0.2750         | 31.61    |
