A Koopmans-compliant screened exchange functional with correct asymptotic behavior for semiconductors

Michael Lorke, Peter Deák, and Thomas Frauenheim

Bremen Center for Computational Materials Science, University of Bremen, Am Fallturm 1, 28359 Bremen, Germany

The performance of density functional theory depends largely on the approximation applied for the exchange functional. We propose here a novel screened exchange potential for semiconductors, with parameters based on the physical properties of the underlying microscopic screening and obeying the requirements for proper asymptotic behavior. We demonstrate that this a functional is Koopmans-compliant and reproduces a wide range of band gaps. We also show, that the only tunable parameter of the functional can be kept constant upon changing the cation or the anion, making the approach suitable for treating alloys.

Density functional theory (DFT) is the workhorse of electronic structure calculations in many areas of solid state physics and materials science. The exact exchange-correlation potential is, unfortunately, not known, and the real predictive power of DFT-based methods is mainly limited by the quality of the approximate exchange functional that they are based upon. It has been established[18] that the exact functional provides a total energy, which is a piecewise linear function of the occupation numbers and has a derivative discontinuity at integer values. This is not case for the standard approximations (LDA: local density approximation, or GGA: generalized gradient approximation). As it is well known, these functionals are convex between, and have no derivative discontinuity between integer occupation numbers. Therefore, they underestimate the gap of semiconductors and lead to an artificial delocalization of defect states. In contrast, DFT with unscreened non-local Hartree-Fock (HF) type exchange leads to a strong overestimation of the band gap and to over-localization of defect states. Earlier, self-interaction correction schemes to LDA/GGA the band gap and over-localization of defect states. A Koopmans-compliant screened exchange potential with correct asymptotic behavior

For example, the PBE0 hybrid[4] is generated by the choice of \(a = 1, \varepsilon^{-1}(q) = 0.25\). If the quasi-momentum dependence of \(\varepsilon^{-1}(q)\) is to be included, a wide range of approaches can be written as

\[
\varepsilon^{-1}(q) = c_1 + c_2 f(q).
\]

Popular choices for \(f(q)\) include a Gaussian,

\[
f(q) = \exp \left( -\frac{q^2}{4\mu^2} \right),
\]

with a screening length \(\mu\), that is used in connection with \(a = 1, c_1 = 0.25\), and \(c_2 = -0.25\) in the popular Heyd-Scuseria-Ernzerhof (HSE) hybrid functionals. Screened exchange methods, such as the sX approximation[6] employ \(a = 0\) and hence no mixing. Specifically, the sX method[6] is defined by \(a = 0, c_1 = 1, c_2 = -1\), and

\[
f(q) = \frac{\kappa^2}{q^2 + \kappa^2},
\]

with a screening length \(\kappa\), corresponding to Thomas-Fermi (TF) theory. However, to get the physically correct limits, \(\varepsilon^{-1}(q \rightarrow 0) = 1/\epsilon_b\), and \(\varepsilon^{-1}(q \rightarrow \infty) = 1\), it is necessary[6] to choose

\[
c_1 = 1, \quad c_2 = \frac{1}{\epsilon_b} - 1.
\]
where $\varepsilon_b$ is the static long-range dielectric constant. These limits ensure that the Coulomb interaction approaches its unscreened value for $r \to 0$ and also the proper $1/r^3$ dependence at $r \to \infty$.

In this work we employ a screened exchange approach, i.e. we use $a = 0$, and screen the HF exchange using an ansatz for $\varepsilon^{-1}(q)$. Correlations are added on the level of a GGA in the PBE approximation\cite{PBE96}. From Green’s function theory, it is known that $\varepsilon^{-1}(q) \propto q^2$ for $q \to 0$, as otherwise the approximation is non-conserving\cite{PBE96}. Via Eq. (2) this directly translates to the same requirement for $f(q)$. Even though Eqs. (3) and (4) fulfill this condition, it is also known\cite{PBE96,TF92} that the $q \to \infty$ limit is not properly described in either approaches. Guided by those works, we employ

$$f(q) = \frac{1}{\cosh(q/\sigma)}$$

(6)

that leads to a quadratic behavior at $q = 0$ and to an exponential decay for $q \to \infty$. To determine the screening length $\sigma$, we connect our approach to TF theory by using

$$\sigma = \frac{2}{2\log(2 + \sqrt{3})}\kappa,$$

(7)

i.e. we choose the same FWHM in our cosh-ansatz as TF theory would give for its Lorentzian of Eq. (4). The screening length $\kappa$ of TF theory can be expressed as\cite{TF92,TF93}

$$\kappa = Z \sqrt{k_{TF}^2 \left( \frac{1}{\varepsilon_b - 1} + 1 \right)},$$

(8)

where an additional renormalization factor $Z$ is introduced, and $k_{TF}$ is the TF wave vector

$$k_{TF} = 4 \left( \frac{3N_{el}}{\pi} \right)^{1/3}.$$

(9)

We will elaborate on the choice of the number of electrons $N_{el}$ per unit cell and on the renormalization factor $Z$ below. Our exchange potential can be seen as a static approximation to a GW calculation with a model screening function. In it’s idea, the method is similar to the sX functional\cite{SSK09}. However, the inclusion of the correct limits and q-dependence marks a substantial improvement. In Fig. 1, we compare our model to the screening used in the HSE and the sX potentials. The sX clearly neglects the background screening $\varepsilon_b$ and uses a Lorentzian q-dependence, while the HSE approach differs substantially and compensates by mixing with PBE exchange.

To apply this theoretical model to semiconductor materials, we have to answer two questions: How to choose $\varepsilon_b$ and $N_{el}$, and what is the meaning of the renormalization factor $Z$? For $\varepsilon_b$ the value resulting from an independent particle calculation of the optical properties on the GGA (PBE) level is used. This is in the spirit of the GW$_0$ approximation\cite{PBE96}, which in most cases reproduces the band gaps better than fully self-consistent GW schemes (without vertex corrections).

It is well known, that energetically deep lying electronic states contribute insignificantly to the screening, as they are strongly localized, their spatial overlap is weak and their energetic distance to the top of the VB is high. Hence we approximate this behavior by disregarding them in the screening completely. The "effective number of electrons", $N_{el}$, is determined as follows. We evaluate which states primarily form the highest valence band and choose the corresponding number of electrons from the constituting atomic states. As an example, in GaN, the top of the VB is made up of nitrogen 2p orbitals. In a nitrogen atom, these hold 3 electrons, hence for the primitive cell containing 2 nitrogen atoms, $N_{el} = 6$. $N_{el}$ should, of course, be increased for a supercell, according to the number of N atoms in it.

The renormalization factor $Z$ has the same meaning as in GW theory\cite{PBE96} and describes the fraction of the main quasi-particle peak of the total frequency dependent spectral function. However, as the value of $Z$ varies for states around the valence band edge, and effects beyond the random phase approximation in the screening could lead to additional modifications, we employ a pragmatic approach, using the GW$_0$ value of $Z$ at the band edge as a starting point, and modify it to reproduce the band gap. As it turns out, in the compounds investigated here, the $Z$ value, which reproduces the band gap and ensures the linearity, lies within ±0.1 of the value from GW$_0$ theory (see below). For these materials, our approach can be regarded as a first approximation to a non-empirical exchange functional that is based on many-body perturbation theory. For other compounds, the value of $Z$ might deviate more. This is currently under study.

We have implemented the exchange functional as discussed above into the Vienna Ab initio Simulation Pack-
age, VASP 5.3, \cite{61,62} using the projector augmented wave method and treating the semi-core d-states as part of the valence shell. The modified Vasp source code can be made available to certified owners of a VASP user license for trial. Calculations on the unit cell of the bulk material where performed using a 6×6×6 Monkhorst-Pack grid. For defect calculations, 64 (GaAs) 96 (GaN), and 160 (Ga2O3) atom supercells were used in the Γ-point approximation. A 450eV (900eV) cutoff was applied for the expansion of the wave functions (charge density). Charge corrections for the total energy were performed by the method of Freysoldt, Neugebauer, and van der Wall\cite{63} while localized defect levels were corrected using the formula derived by Chen and Pasquarello\cite{64}. GW gaps were calculated on top of PBE calculations with 1000 bands and a 6×6×6 Γ-centered Monkhorst-Pack grid.

For the purpose of this letter, we focus on the technologically important gallium based compounds GaAs, GaN, and Ga2O3, as well as the ternary compounds CuGaS2 and CuInS2. On the one hand, the gaps of these semiconductors span a range between 1.5 and 5.0 eV, and are systematically underestimated by standard HSE06\cite{65}. On the other hand, these compounds represent different lattice structures and vary in composition and in the degree of ionicity. Therefore, they can serve as a representative set for testing the method.

Table I shows the calculated band gaps, in comparison with the GW0 results. The dielectric constant calculated by PBE, the appropriate value of the effective number of electrons, as well as the optimal value of the scaling factor Z is also given. Excellent agreement of the band gap with the GW0 values is found for all materials under investigation. We want to note that our value for the quasi-particle band gap, as given in Table I is determined to include neither excitonic nor polaronic renormalizations, as both the electron-hole Coulomb interaction and the electron-phonon interaction lower the band gap (resulting in the measurable optical gap) and are not included in our theory\cite{20,21}.

| Material   | $E_{\text{target}}^\text{G (GW0)}$ | $N_{\text{el}}$ | Z     | $\varepsilon_b$ | $E_{\text{G(present work)}}$ |
|------------|----------------------------------|----------------|-------|-----------------|-----------------------------|
| GaAs       | 1.5\textsuperscript{a}            | 3              | 0.72  | 24.1            | 1.59                        |
| GaN        | 3.6\textsuperscript{a}            | 6              | 0.72  | 6.1             | 3.62                        |
| Ga2O3      | 5.4\textsuperscript{a}            | 24             | 0.72  | 3.93            | 5.08                        |
| CuGaS2     | 2.6                               | 16             | 0.65  | 10.5            | 2.6                         |
| CuInS2     | 1.5                               | 16             | 0.65  | 10.8\textsuperscript{b} | 1.5                         |

\textsuperscript{a} GW0 calculation as described above
\textsuperscript{b} As CuInS2 is metallic in PBE, so we use the same difference in $\varepsilon_b$ to CuGaS2 as in their experimental values.

As discussed above, the optimal renormalization factor of our approach is close to its value at the valence band maximum of the corresponding GW0 calculation, e.g. Z = 0.78 for GaN and Z = 0.66 for CuGaS2. The deviations most probably stems from the fact, that a static approximation to the GW selfenergy is employed and hence they reflect the influence of frequency dependent screening. Our result for the optimal Z value means that our treatment can handle semiconductors of a similar type (cation or anion replaced) with a common parameter Z, and hence the method is well suited to handle alloys of isovalent elements. We want to point out, that in an alloy $\varepsilon_b$ has to be recalculated for the given composition, while $N_{\text{el}}$ is determined by the prevailing crystal structure.

As discussed in previous works\cite{20,21}, not only the minimum band gap, but the band edges positions over the entire Brillouin-zone have to fit the results of a GW0 calculation, in order to provide a proper description of defect levels, as they can be generated by a superposition of all band edge states. This condition is satisfied in all the compounds of Table I as shown on the example of GaN in Fig. 2.

\begin{table}[h]
\begin{tabular}{|c|c|c|c|}
\hline
Material & $\Delta KS (\text{HOMO}) - \Delta SCF$ & $\Delta SCF - \Delta KS (\text{LUMO})$ \\
\hline
GaAs     & 0.02                              & 0.03                        \\
GaN      & -0.04                             & -0.04                      \\
Ga2O3    & -0.03                             & -0.04                      \\
\hline
\end{tabular}
\caption{Fullfillment of the gKT for various materials. $\Delta KS (\text{HOMO})$ and $\Delta KS (\text{LUMO})$ describe the energetic position of the Kohn-Sham levels of the highest occupied molecular orbital for the neutral (HOMO), and of the lowest unoccupied molecular orbital (LUMO) for the +1 charge state of the vacancy. $\Delta SCF$ is the electron removal energy.}
\end{table}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{bandstructure.png}
\caption{Band structure of GaN, calculated with the GW0 approach (lines) as well as with the screened exchange approach presented in this work (squares). The energy of the valence band maximum has been set to 0.}
\end{figure}

To correctly describe the energetic position and the lo-
calization of defect states, the gKT has to be fulfilled\footnote{\ref{13,14}} i.e., the total energy has to be linear with respect to the fractional occupation number. We have tested this criterion on the (0/+1) transition levels of the antisite pair in GaAs, the nitrogen vacancy in GaN, and the oxygen vacancy in Ga$_2$O$_3$. As shown in Table \ref{table1}, we find an excellent fulfillment of the gKT.

In conclusion, we presented here a novel exchange functional for semiconductors. As it is based on the correct limits of the exchange potential, its parameters can be derived from physical principles, to a large extent eliminating the need for "tuning" which is a common procedure with current hybrid functionals. Because of the correct piecewise linearity of the total energy, this functional reproduces the relative position of the band edge states and fulfills the generalized Koopmans' theorem. This is of high significance in various application areas. For example, in studying photo-assisted reactions on semiconducting catalysts, it is critical to reproduce the position of the band edges, which measure the chemical potential of photo-generated holes and electrons\footnote{\ref{15-17}}. In general, the correct gap is the starting point to determine the optical and transport properties as well. Besides the reproduction of the gap, the fulfillment of the gKT is a condition to accurately predict the localization and energy of defects states in semiconductors, which is a prerequisite in the successful identification of defects, which influence device behavior in micro/opto-electronics and photovoltaics\footnote{\ref{18-21}}. A great advantage of our approach is that it can be used to describe alloys without any re-tuning, as the single parameter of this approach is transferable between semiconductors of a similar type, when the cation or anion is replaced.

\section*{Acknowledgments}

The authors would like to acknowledge stimulating discussions with G. Kresse. Funding from the DFG via research project FR2833/63-1 and the support of the Supercomputer Center of Northern Germany via HLRN Grant No. hbc00027 is acknowledged.

\begin{thebibliography}{99}
\footnotesize
\bibitem{1} John P. Perdew, Robert G. Parr, Mel Levy, and Jose L. Balduz. Density-functional theory for fractional particle number: Derivative discontinuities of the energy. \textit{Phys. Rev. Lett.}, 49:1691–1694, Dec 1982.
\bibitem{2} J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. \textit{Phys. Rev. B}, 23:5048–5079, May 1981.
\bibitem{3} Björn Baumeier, Peter Krüger, and Johannes Pollmann. Self-interaction-corrected pseudopotentials for silicon carbide. \textit{Phys. Rev. B}, 73:195205, May 2006.
\bibitem{4} Stewart J. Clark and John Robertson. Screened exchange density functional applied to solids. \textit{Phys. Rev. B}, 82:085208, Aug 2010.
\bibitem{5} S. J. Clark, J. Robertson, S. Lany, and A. Zunger. Intrinsic defects in zno calculated by screened exchange and hybrid density functionals. \textit{Phys. Rev. B}, 81:115311, Mar 2010.
\bibitem{6} Axel D. Becke. Density functional thermochemistry, iv. a new dynamical correlation functional and implications for exact-exchange mixing. \textit{The Journal of Chemical Physics}, 104(3):1040–1046, 1996.
\bibitem{7} C. Adamo and V. Barone. \textit{J. Chem. Phys.}, 110:6158, 1999.
\bibitem{8} J. Heyd, G. E. Scuseria, and M. Ernzerhof. \textit{J. Chem. Phys.}, 118:8207, 2003.
\bibitem{9} Aliaksandr V. Krukau, Gustavo E. Scuseria, John P. Perdew, and Andreas Savin. Hybrid functionals with local range separation. \textit{The Journal of Chemical Physics}, 129(12):124103, 2008.
\bibitem{10} Xiao Zheng, Aron J. Cohen, Paula Mori-Sánchez, Xiangqian Hu, and Weitao Yang. Improving band gap prediction in density functional theory from molecules to solids. \textit{Phys. Rev. Lett.}, 107:026403, Jul 2011.
\bibitem{11} Andrius Alkauskas, Peter Broqvist, and Alfredo Pasquarello. Defect levels through hybrid density functionals: Insights and applications. \textit{physica status solidi (b)}, 248(4):775–789.
\bibitem{12} Wei Chen, Giacomo Miceli, Gian-Marco Rignanese, and Alfredo Pasquarello. Nonempirical dielectric-dependent hybrid functional with range separation for semiconductors and insulators. \textit{Phys. Rev. Materials}, 2:073803, Jul 2018.
\bibitem{13} Peter Deák, Michael Lorke, Bálint Aradi, and Thomas Frauenheim. Optimized hybrid functionals for defect calculations. \textit{submitted}.
\bibitem{14} J. F. Janak. Proof that $\frac{\partial E}{\partial n} = \epsilon$ in density-functional theory. \textit{Phys. Rev. B}, 18:7165–7168, Dec 1978.
\bibitem{15} Mel Levy, Rajeev K. Pathak, John P. Perdew, and Siqing Wei. Indirect-path methods for atomic and molecular energies, and new koopmans theorems. \textit{Phys. Rev. A}, 36:2491–2494, Sep 1987.
\bibitem{16} Stephan Lany and Alex Zunger. Polaronic hole localization and multiple hole binding of acceptors in oxide wide-gap semiconductors. \textit{Phys. Rev. B}, 80:085202, Aug 2009.
\bibitem{17} John P. Perdew and Mel Levy. Comment on “significance of the highest occupied kohn-sham eigenvalue”. \textit{Phys. Rev. B}, 56:16021–16028, Dec 1997.
\bibitem{18} Peter Deák, Bálint Aradi, Thomas Frauenheim, Erik Janzén, and Adam Gali. Accurate defect levels obtained from the hse06 range-separated hybrid functional. \textit{Phys. Rev. B}, 81:153203, Apr 2010.
\bibitem{19} Miaomiao Han, Zhi Zeng, Thomas Frauenheim, and Peter Deák. Defect physics in intermediate-band materials: Insights from an optimized hybrid functional. \textit{Phys. Rev. B}, 96:165204, Oct 2017.
\bibitem{20} Peter Deák, Quoc Duy Ho, Florian Seemann, Bálint Aradi, Michael Lorke, and Thomas Frauenheim. Choosing the correct hybrid for defect calculations: A case study on intrinsic carrier trapping in $\beta - \text{Ga}_2\text{O}_3$. \textit{Phys. Rev. B}, 95:075208, Feb 2017.
\bibitem{21} Peter Deák, Michael Lorke, Bálint Aradi, and Thomas Frauenheim. Carbon in gan: Calculations with an opti-
mized hybrid functional. *Phys. Rev. B*, 99:085206, Feb 2019.

22 J. P. Perdew, K. Burke, and M. Ernzerhof. *Phys. Rev. Lett.*, 77:3865, 1996.

23 P. Gartner, L. Bányai, and H. Haug. Coulomb screening in the two-time Keldysh-Green-function formalism. *Phys. Rev. B*, 62:7116, 2000.

24 L. Bányai, P. Gartner, and H. Haug. Self-consistent RPA retarded polaron Green function for quantum kinetics. *Eur. Phys. J. B*, 1:209, 1998.

25 P. Gartner, L. Bányai, and H. Haug. Self-consistent RPA for the intermediate-coupling polaron. *Phys. Rev. B*, 66:75205, 2002.

26 G. Cappellini, R. Del Sole, Lucia Reining, and F. Bechstedt. Model dielectric function for semiconductors. *Phys. Rev. B*, 47:9892–9895, Apr 1993.

27 Tomomi Shimazaki and Yoshihiro Asai. Band structure calculations based on screened fock exchange method. *Chemical Physics Letters*, 46(1):91 – 94, 2008.

28 Tomomi Shimazaki and Yoshihiro Asai. Energy band structure calculations based on screened hartree–fock exchange method: Si, aip, alas, gap, and gaas. *The Journal of Chemical Physics*, 132(22):224105, 2010.

29 M. Shishkin and G. Kresse. Self-consistent GW calculations for semiconductors and insulators. *Phys. Rev. B*, 75:235102, Jun 2007.

30 technically from VASP’s POTCAR file.

31 G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mat. Sci.*, 6(1):15–50, July 1996.

32 G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169, 1996.

33 Hendrik J. Monkhorst and James D. Pack. Special points for brillouin-zone integrations. *Phys. Rev. B*, 13:5188–5192, Jun 1976.

34 Christoph Freysoldt, Jörg Neugebauer, and Chris G. Van de Walle. Fully ab initio finite-size corrections for charged-defect supercell calculations. *Phys. Rev. Lett.*, 102:016402, Jan 2009.

35 Wei Chen and Alfredo Pasquarello. Correspondence of defect energy levels in hybrid density functional theory and many-body perturbation theory. *Phys. Rev. B*, 88:115104, Sep 2013.

36 J. Furthmüller and F. Bechstedt. Quasiparticle bands and spectra of ga₂o₃ polymorphs. *Phys. Rev. B*, 93:115204, Mar 2016.

37 Wei-Na Zhao and Zhi-Pan Liu. Mechanism and active site of photocatalytic water splitting on titania in aqueous surroundings. *Chem. Sci.*, 5:2256–2264, 2014.

38 Peter Deák, Jolla Kullgren, Beálint Aradi, Thomas Frauenheim, and Ladislaw Kavan. Water splitting and the band edge positions of TiO₂. *Electrochimica Acta*, 199:27 – 34, 2016.