Hybrid functionals with fixed mixing parameter perform no better than PBE for fundamental band gaps of nanoscale materials

Xinquan Wang, Marc Dvorak, and Zhigang Wu
DOI: 10.1103/PhysRevB.94.195429
Hybrid functionals with fixed mixing parameter perform no better than PBE for fundamental band gaps of nanoscale materials

Xinquan Wang, Marc Dvorac, and Zhigang Wu

Department of Physics, Colorado School of Mines, Golden, CO 80401, USA

(Dated: October 28, 2016)

Hybrid functionals mixing the exact exchange with (semi)local functionals to reinstall the missing derivative discontinuity, have been successfully employed to predict band gaps ($E_g$) in bulk semiconductors. Here we show that traditional hybrid functionals with fixed fractions of exact exchange do not perform significantly better than the most popular semilocal PBE-GGA functional for $E_g$ of semiconductor nanostructures, since their band-gap corrections are essentially size independent. This is because they cannot respond properly to the variation in screening when size changes. They merely predict constant band gap corrections to the PBE gaps in silicon nanowires (Si NWs) when wire diameter reduces, instead of the dramatic increase predicted by many-body $GW$ calculations. Moreover, these hybrid functionals generate almost identical wave functions compared with PBE for both bulk Si and Si NWs, whose overlaps with corresponding quasiparticle wave functions become much smaller than 1 for narrow NWs.

PACS numbers: 73.21.Hb, 73.22.Dj, 78.67.Lt

I. INTRODUCTION

The Kohn-Sham (KS) formulation of density functional theory (DFT)\textsuperscript{1,2} has become the dominating method for predicting the ground-state properties of condensed matter systems, due largely to the simplicity of the exchange-correlation (XC) energy in either local density (LDA) or generalized gradient (GGA) approximations. However, lack of derivative discontinuity with respect to the electron number\textsuperscript{3–5} within these (semi)local approximations leads to severe underestimation of fundamental band gaps ($E_g$).\textsuperscript{6–9} This problem can be resolved in the framework of the many-body perturbation theory (MBPT)\textsuperscript{10,11} by solving the quasiparticle (QP) equation.\textsuperscript{12} A practical scheme is the $GW$ approximation\textsuperscript{6–9}, within which the self-energy ($\Sigma$) is obtained from the Green function ($G$) and the dynamically screened Coulomb interaction ($W$) by

$$\Sigma = iGW.$$

The $GW$ method usually predicts reasonable band gaps for a broad range of semiconductors and insulators\textsuperscript{13–17}, but computationally it is extremely demanding, especially for nanoscale materials.

An alternative remedy lies within the generalized KS (GKS) scheme\textsuperscript{18–20}, which reinstall the missing derivative discontinuity using a single Slater determinant. The GKS equation is written as:

$$\left(\hat{O}_s[\psi_i] + V_{\text{ext}}(r) + V_{\text{R}}(r)\right)\psi_j(r) = \varepsilon_j\psi_j(r),$$

where $\hat{O}_s[\psi_i]$ is a nonlocal orbital-dependent operator depending on the choice of the Slater determinant $S[\{\psi_i\}]$, $V_{\text{ext}}(r)$ is the external potential, and $V_{\text{R}}(r)$ is the remainder potential including XC, Hartree, and kinetic energy components not accounted for by $\hat{O}_s[\psi_i]$. It is generally hoped that the GKS band gap ($E_g^{\text{GKS}}$) produced by a judiciously chosen $\hat{O}_s[\psi_i]$ will be considerably closer to the QP band gap than $E_g^{\text{KS}}$, because $\hat{O}_s[\psi_i]$ inherently exhibits the derivative discontinuity due to its orbital dependence\textsuperscript{18,19,21,22}. It is further expected that the nonlocal character of $\hat{O}_s[\psi_i]$ would allow it to efficiently mimic the role played by $\Sigma$ in MBPT calculations of QP energies\textsuperscript{23,24}.

The hybrid functional approach is a special case of GKS\textsuperscript{18}, in which $\hat{O}_s[\psi_i]$ corresponds to the sum of single-particle kinetic energy operator and an empirically weighted mixture of approximate (semi)local exchange and orbital-dependent exact Fock exchange. Popular hybrid functionals such as B3LYP\textsuperscript{25}, PBE0\textsuperscript{26}, and HSE\textsuperscript{27}, can predict good band gaps for bulk materials\textsuperscript{28,29}, and thus they are considered as an interesting compromise between KS and $GW$ methods, because of their considerable accuracy and relatively low computational cost. Hybrids have been widely employed to investigate electronic structures of a variety of materials, including nanostructures\textsuperscript{31–34}, however, very few efforts\textsuperscript{30} have been spent on justifying and examining their reliability for nanoscale systems.

In this work, we demonstrate that hybrid functionals with fixed fraction of exact exchange are not reliable to predict band gaps for nanoscale materials at all. Even if a hybrid functional could give a bulk band gap close to the experimental value, this functional fails to improve $E_g$ over simple (semi)local functionals for low-dimension materials with strong quantum confinement. In particular, as diameter of a silicon nanowire (Si NW) reduces, hybrid functionals yield almost constant band gap corrections, instead of the dramatic size-dependent variation correctly captured by the $GW$ approximation. They also fail to improve the poor quality of the KS wave functions (WFs) for the conduction states in Si NWs, thus we conclude that for electronic structure calculations these traditional hybrid functionals are no better than (semi)local functionals, and their success in simple semiconductors is
largely fortuitous.

II. FORMALISM AND COMPUTATIONAL METHODS

The simplest formula of the XC energy, $E_{xc}^{hyb}$, of a hybrid functional can be expressed as:

$$E_{xc}^{hyb} = bE_{x}^{HF} + (1 - b)E_{xc}^{LDA/GGA} + E_{c}^{LDA/GGA},$$

(3)

where the parameter $b$ indicates the portion of exact Hartree-Fock (HF) exchange, $E_{x}^{HF}$, while $E_{xc}^{LDA/GGA}$ and $E_{c}^{LDA/GGA}$ respectively denote the exchange and correlation energy for the (semi)local functional. Hybrid functionals with different portions of exact exchange could give the same ground-state total energy and density, with different eigenvalues $\varepsilon_j$ and eigenfunctions $\psi_j(r)^{18}$.

Specifically, PBE0$^{26}$ was constructed by simply mixing 25% exact exchange with 75% of PBE (Perdew-Burke-Ernzerhof formulation of GGA$^{45}$) exchange, while B3LYP$^{25}$ is a three-parameter hybrid functional:

$$E_{xc}^{B3LYP} = 0.20E_{x}^{HF} + 0.80E_{xc}^{LDA} + 0.72(E_{xc}^{B88} - E_{xc}^{LDA}) + 0.81E_{c}^{LDA} + 0.19E_{c}^{LDA},$$

(4)

where $E_{xc}^{B88}$ and $E_{c}^{LDA}$ are the Becke 88 [Ref. 36] and Lee-Yang-Parr$^{37}$ exchange functionals, respectively. HSE is a range-separated hybrid functional, which splits the exchange term into short-range (SR) and long-range (LR) parts to avoid the demanding integrals over the long-range part of exact exchange:

$$E_{xc}^{HSE} = 0.25E_{x}^{SR} + 0.75E_{xc}^{PBE} + E_{xc}^{PBE} + E_{c}^{PBE}.$$

(5)

Though HSE contains the same portions of exact exchange and PBE exchange as PBE0 in its short-range exchange part, excluding the long-range part of exact exchange causes a significantly different effective screening environment for HSE relative to PBE0$^{27}$.

To put our verification regarding the reliability of hybrid functionals for nanoscale materials in concrete terms, we choose silicon nanowires (Si NWs$^{38,39}$), one of the most important semiconductor nanomaterials, to study the gap scaling law as a function of quantum confinement. Si NWs are the prototypical one-dimensional (1D) nanostructures, exhibiting both 1D extension along the wire axis and 2D confinement perpendicular to it. Previous calculations$^{40-42}$ show that LDA and GGAs not only underestimate $E_g$ but also lead to an increasing error in $E_g$ as wire diameter ($d$) decreases. Thus one cannot use a single scissor shift to correct KS band gaps for Si NWs with varying size. Here, we examine whether hybrid functionals can lead to meaningful improvement over KS in predicting $E_g$.

We investigated hydrogen-passivated Si NWs grown along (100), (110) and (111) directions, with $d$ up to 4.0 nm. All cylindrical wire models were constructed without SiH$_3$ on the surface. We performed structural relaxations using the plane-wave VASP package$^{43}$ using the PBE functional and adopting the projector augmentation wave method$^{45}$. 10 Å of vacuum was added between NWs, and a plane-wave cutoff of 400 eV and 16 k-points along the wire axis were used to guarantee convergence within 0.01 eV for the calculated band gaps. The hybrid functional calculations were also carried out using the VASP package and the PBE pseudopotentials. Such inconsistency in pseudopotential only leads to a negligible error ($\leq 0.05$ eV) in band gap for Si-based materials$^{39}$.

Since accurately measuring band gaps of Si NWs remains a major challenge and so far barely any experimental data$^{39}$ are available, we examine hybrid functional results by comparing with the benchmarks set by the highly accurate GW calculations, which have been successfully carried out for Si NWs$^{41,42,47,48}$ using pseudopotentials. Our one-shot $G_0W_0$ calculations were performed using the ABINIT code$^{49}$ with pseudopotentials. To overcome the long-range image interactions for NWs due to the periodic boundary condition, we employed a cylinder truncation scheme$^{42,47,48,50}$, in which the truncated Coulomb interaction for a wire oriented along the z-axis is given by

$$V_c(r) = \frac{e^2}{|r|}\theta(\rho - \rho_c),$$

(6)

where $\rho$ is the cylindrical coordinate perpendicular to the wire axis $z$, and $\rho_c$ is the radial cutoff and has been set to half of the distance between the neighboring wires in our calculations. We adopted two plasmon-pole models (PPMs) of Hybertsen-Louie (HL$^{51}$) and Godby-Needs (GN$^{52,53}$) to approximate the frequency dependence of the dielectric function$^{10}$ ($\varepsilon$), and previous studies$^{54,59}$ have found that $E_g$ obtained by numerically explicit integration using the contour-deformation approach lay between the results of HL and GN. We used PBE wavefunctions and eigenenergies as input for evaluating the self-energy, employed the norm-conserving pseudopotentials, and set the energy cutoffs for the ground-state and the GW-kernel calculations to be 14 Ha. We used the extrapolation technique proposed by Bruneval and Gonze$^{57}$ and we found that convergence was achieved when the number of conduction bands was set to twice of that for valence bands in each Si NW, using 16 k-points along the wire axis.

III. RESULTS AND DISCUSSION

Our calculated $E_g$ for bulk Si are 1.69, 1.71, 1.20 and 6.62 eV, using PBE0, B3LYP, HSE and HF, respectively, agreeing well with previous benchmark calculations$^{46}$. Compared with PBE value of 0.59 eV, hybrid functionals considerably exaggerate the bulk band gap. B3LYP and PBE0 essentially predict the same band gap, much larger than the corresponding HSE value. This is because of the range-separation approach adopted in HSE, where the exact exchange is short ranged and consequently leads to the much reduced $E_g$ compared to PBE0 and B3LYP.
The present GW band gaps for bulk Si are 1.27 and 1.22 eV using HL and GN PPMs, respectively, in good agreement with previous works\cite{54,56,58}. The GW band gaps for Si NWs are plotted in Fig. 1, which clearly shows that the two PPMs produce very close $E_g$ for a given Si NW. Despite the similarity, the GN band gap is always slightly smaller than that of HL, and the difference in $E_g$ between HL and GN gradually increases from 0.05 eV for bulk Si to about 0.2 eV for a 1nm NW. Recent works have revealed that GN and HL PPMs could differ significantly from each other for strongly localized electronic systems, but they behave similarly for highly delocalized systems\cite{54,56}, because HL enforces the fulfillment of the $f$-sum rule which tends to underestimate the screening and consequently produce larger band gaps\cite{56,59}. In this case, as $d$ shrinks the electron localization in NWs increases, so that HL tends to overestimate $E_g$ more.

Because of strong quantum confinement in narrow Si NWs, band gaps can be well described by the effective-mass approximation (EMA) model\cite{40,60},

$$E_{g,NW} = E_{g,bulk} + \frac{C}{d^{\alpha}},$$

where the exponent $\alpha$ is equal to 2 within EMA\cite{64}, and
the confinement constant \( C \) can be expressed as \( C = 2\hbar^2\zeta^2/m^* \), with \( \zeta = 2.4048 \) the first root of the zeroth-order cylindrical Bessel function of the first kind and \( m^* \) the effective mass. As shown in Fig. 1, all these methods produce \( E_g \) for Si NWs satisfying the scaling law (Eq. 6); but they predict different values of parameters \( C \) and \( \alpha \). The KS scheme underestimates not only \( E_g \) in Si NWs, but also the trend, i.e., values of \( C \) and \( \alpha \) are also underestimated, compared to the GW results.

Table I summarizes the fitted values for parameters \( \alpha \) and \( C \). GW predicts \( \alpha \) to be around 1.4–1.5, smaller than the ideal value of 2 because the EMA assumes an unrealistic infinite potential at the wire surface. Another observation is that the constant \( C \) for \((100)\) and \((111)\) wires are rather close, but more than 50\% larger than that for the \((110)\) wires. Such anisotropy reflects the orientation dependence of the Si NW band gaps caused by different effective mass \( m^* \) in the confinement plane:

\[
m_{(100)}^* \sim m_{(111)}^* < m_{(110)}^*.
\]

PBE predicts slightly smaller \( \alpha \) but severely underestimates \( C \) compared with GW, leading to a fast increasing error in \( E_g \) as \( d \) reduces. Although HF exaggerates band gaps by 6 eV, its fitted \( \alpha \) and \( C \) are the smallest among all these methods; in particular, its \( \alpha \) are very close to 1.0 while \( C \) are slightly smaller than the PBE values except for the \((100)\) wires. The confinement constant \( C \) is inversely proportional to the effective mass \( m^* \), which is determined from the dispersion relation at the band edge by

\[
m^* = \frac{\hbar^2}{\partial^2 E/\partial^2 k_z}.
\]

Our band structure analysis for Si NWs confirms the trend in \( m^* \) obtained by using various functionals.

Next we focus on \( \alpha \) and \( C \) obtained from hybrid functionals and examine if they can predict better trends. Surprisingly, these three hybrid functionals give \( \alpha \) and \( C \) comparable to those of PBE. Among them HSE almost reproduces the PBE values, while PBE0 and B3LYP predict smaller \( \alpha \) and larger \( C \) than PBE. Previously researchers mostly focused on exponent \( \alpha \) and largely ignored the role played by the confinement constant \( C \) in the band gap scaling; it was expected that better density functional would principally improve \( \alpha \). However, Table I demonstrates that the major issue of the (semi)local functionals lies in severe underestimation of \( C \) (the ratios of \( C_{GW}/C_{PBE} \) are around 1.6–1.7 for \((100)\) and \((111)\) wires while \( \sim 2.0 \) for the \((110)\) NWs, respectively) instead of \( \alpha \) (\( \alpha_{GW}/\alpha_{PBE} \) is in the range of 1.05–1.15). In this sense, B3LYP and PBE0 perform even worse than PBE, while HSE behaves very similar to PBE.

The hybrid functionals band-gap corrections \((\Delta E_{\text{hyb}}^{\text{GW}} = E_{\text{hyb}}^{\text{GW}} - E_{\text{PBE}}^{\text{hyb}})\) to PBE are in general just constants as functions of \( d \); \( \Delta E_{\text{hyb}}^{\text{GW}}(d) \) even decreases when \( d \) decreases towards 1 nm, while the GW corrections \( \Delta E_{\text{hyb}}^{\text{GW}} \propto 1/d^\alpha \) (Fig. 2), increasing dramatically. In fact, \( \Delta E_{\text{hyb}}^{\text{GW}}(d) \approx \Delta E_{\text{hyb}}^{\text{GW, bulk}} \), i.e., the hybrid functional band-gap corrections for any Si NWs are essentially the same as those for bulk Si.

Previously a size-independent constant correction \((\sim 0.6\) eV from bulk Si) to the KS band gaps of Si NWs was proposed to study their optical properties,\(^62\) omitting the excitonic binding; however, the quasiparticle correction could partially cancel the excitonic correction.\(^61\) Compared to this scissorshift method, hybrid functionals offer no further improvement over PBE on band structures for these nanoscale materials.

Within the GW approximation, self-energy is a nonlocal frequency-dependent operator sensitive to the screening environment. The real part of \( \Sigma \) could be split into two components: \( \Re(\Sigma) = \Re(\Sigma_{\text{SEX}}) + \Re(\Sigma_{\text{COH}}) \), where the local \( \Sigma_{\text{COH}} \) represents the Coulomb-hole (COH) interaction and the nonlocal \( \Sigma_{\text{SEX}} \) is the statically screened exchange (SEX):

\[
\Sigma_{\text{COH}}(\mathbf{r}, \mathbf{r}') = -\frac{1}{2} \delta(\mathbf{r} - \mathbf{r}') \left[ v(\mathbf{r}, \mathbf{r}') - W(\mathbf{r}, \mathbf{r}') \right],
\]

\[
\Sigma_{\text{SEX}}(\mathbf{r}, \mathbf{r}') = -\sum_{i=1}^{N_{\text{occ}}} \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}') W(\mathbf{r}, \mathbf{r}') ,
\]

in which \( v \) is the bare Coulomb potential, and \( N_{\text{occ}} \) represents the number of occupied bands. As WFs become more confined and localized, \( \Sigma_{\text{SEX}} \) is gradually enhanced since the screening \( (W) \) becomes weaker,\(^47,76\) and then dominates the self-energy and the GW band-gap corrections consequently show a strong size dependence. But the hybrid functional \( E_{\text{hyb}}^{\text{GW}} \) cannot respond the change in screening as well as the self-energy operator \( \Sigma \). This is because the fraction of the exact exchange is fixed, so that the average dielectric screening that a hybrid functional offers varies only due to the (semi)local functional part; therefore it behaves just like these LDA/GGA functionals when confinement strength changes except for a constant band-gap shift. For simple bulk materials such as Si, the valence electrons are very delocalized,\(^54,55\) and the screening could be similar among these materials. As a result, although a well constructed hybrid functional, such as HSE (a screened PBE hybrid), can well predict \( E_g \) of these materials, the success is more or less fortuitous and empirical due to cancellation of overestimation of \( E_g \) by HF and underestimation by LDA/GGA. For
We performed fully quasiparticle self-consistent LDA/GGA WFs can still well represent QP WFs in NWs. The distribution for the electron density (GECO) at the CSM and National Energy Research Scientific Computing Center (NERSC).

materials with sharply different valence electron localization and screening compared to simple semiconductors, such as transition metal oxides, errors in $E_g$ will vary dramatically, and hybrid functionals with a fixed mixing parameter can not work equally well for both cases. Here we show that even for the same material (Si) as the size approaches the quantum confinement regime, hybrid functionals with fixed fractions of exact exchange fail completely to predict the size-dependent band-gap correction as GW does.

Though the (semi)local functionals severely underestimate $E_g$, even LDA WFs can provide excellent approximations to the QP WFs for simple bulk systems, with overlaps $|\langle \psi_{\text{LDA}} | \psi_{\text{QP}} \rangle | \geq 0.999$. It is not clear if LDA/GGA WFs can still well represent QP WFs in NWs. We performed fully quasiparticle self-consistent GW calculations in which the WFs are updated together with $G$ and $W$ until convergence reached $13,71$. Because the self-consistent GW computation is extremely expensive, we only carried out such calculations for a (100) 0.71 nm Si NW ($\text{Si}_8\text{H}_{12}$). Fig. 3 shows electron density $|\langle \psi (r) \rangle^2|$ at the $\Gamma$-point for the conduction band minimum (CBM) and the CBM+4 state, in comparison with the second and fourth states below valence band maximum (VBM-2 and VBM-4). These states are chosen to avoid degeneracy. Surprisingly, even in such a narrow wire the PBE WFs for valence bands still overlap with QP ones nearly perfectly $|\langle \psi_{\text{PBE}} | \psi_{\text{QP}} \rangle | \approx 0.996$; however, the quality of the KS WFs for conduction bands is much poorer, with overlaps only about 81.9%, 42.0%, 18.5% and 39.2% for the CBM, CBM+3, CBM+4, CBM+5 states, respectively. Hybrid functionals cannot improve WFs over PBE at all. For all these states of the 0.71 nm Si NW we find that $|\langle \psi_{\text{PBE}} | \psi_{\text{hyb}} \rangle | \geq 0.998$.

Although our conclusion on the performance of hybrid functionals for nanoscale materials is based on calculations on Si NWs alone, we believe that it holds for other semiconductor NWs and nanocrystals, as previous calculations revealed very similar results for band gap scaling to those of Si NWs. Finally we would like to point out that in this work we have focused on traditional and widely employed hybrid functionals such as HSE, PBE0 and B3LYP. We have noticed that recent developments in hybrid functionals proposed setting the mixing parameter $b = 1/\epsilon$, in order to respond better to the change in dielectric environment. These new hybrid functionals thus could predict good trend of $E_g$ in nanoscale semiconductors, though computationally they are more demanding than those hybrid functionals with fixed $b$.

IV. SUMMARY

In conclusion, we have investigated the accuracy of hybrid functionals with fixed mixing parameters for band gaps of nanoscale materials. Comparing with the GW benchmarks, we find that these hybrid functionals only yield almost constant gap corrections to PBE results as quantum confinement increases, instead of reproducing the crucial size-dependent GW corrections, though in general the calculated band gaps by hybrid functionals are closer to GW values than LDA or GGA. Hybrid and (semi)local functionals obtain very similar parameters for the band-gap scaling law as a function of wire diameter $d$, both severely underestimating the confinement constant $C$ because they are not able to response the change in screening properly. Furthermore, WFs obtained using hybrid functionals are very similar to those of (semi)local functionals even in strongly confined NWs. Therefore, these hybrid functionals have limited applicability for investigating electronic structures of novel materials, especially for those with more localized electrons than bulk semiconductors. However, recently developed hybrid functionals with mixing parameters according to material’s dielectric constant could solve this problem.

V. ACKNOWLEDGEMENTS

This work was financially supported by DOE Early Career Award (No. DE-SC0006433). Computations were carried out at the Golden Energy Computing Organization (GECO) at the CSM and National Energy Research Scientific Computing Center (NERSC).
1. L. Hedin, Phys. Rev. A