2-1

New approaches for first-principles modelling of inelastic transport in nanoscale semiconductor devices with thousands of atoms

Tue Gunst*, Mads Brandbyge*, Mattias Palsgaard*†, Troels Markussen† and Kurt Stokbro†
*DTU Nanotech Center for Nanostructured Graphene (CNG), Technical University of Denmark, DK-2800 Kgs. Lyngby Denmark.
†QuantumWise A/S, Fruebjergvej 3, DK-2100 Copenhagen, Denmark. Email:kurt.stokbro@quantumwise.com

***Abstract*—We present two different methods which both enable large-scale first-principles device simulations including electron-phonon coupling (EPC). The methods are based on Density Functional Theory and Nonequilibrium Greens Functions (DFT-NEGF) calculations of electron transport. The inelastic current is in both methods calculated in a post-processing step to a self consistent DFT calculation. The first method is based on first order perturbation theory in the EPC self-energy within the Lowest Order Expansion (LOE) approximation. The method requires calculation of the first-principles EPC in the device region and it includes the effect of each phonon mode on the current perturbatively. This approach is made practical by calculating the EPC of the device region using a smaller periodic reference system. In addition, the phonon modes are assembled into a small number of energy intervals in which phonon modes are described collectively. The second method involves calculating the electron transmission for a single configuration where the atoms are displaced according to the phonon temperature of the system. Thus, this method has a computational cost equivalent to conventional elastic transport calculations. Both methods have been implemented in the Atomistix ToolKit (ATK) and we apply the methods for calculating the inelastic current in a silicon n-i-n junction and for calculation of phonon limited mobilities of silicon nanowires.**

I. Introduction

As electronic devices approach the nanoscale, accurate modelling often requires that the effect of each individual atom is included in the simulation. First-principles modelling using DFT-NEGF is an attractive approach, since it can accurately describe the atomic-scale details and electronic structure of surfaces, interfaces and different material combinations without the use of any experimental data[1]. However, most studies have so far been limited to simulations of the elastic current, even though EPC is known to play a crucial role in room-temperature performance of many nanoscale devices[2]. The inclusion of EPC into first-principles transport calculations has so far been limited to small molecular systems, due to the high computational cost[3]. In this paper, we present two methods[4] which enable electron transport calculations with EPC at a computational cost similar to that of elastic quantum transport calculations. Both methods are implemented in the Atomistix ToolKit (ATK)[5].

In the first section we present the basic theory behind the methods, and in the second section we compare the methods for calculating the inelastic transport in a 2D n-i-n silicon device and for the phonon limited mobility of a silicon nanowire.

II. Theory

In this section we briefly introduce the methods for calculating the inelastic current for a nanoscale device. We will assume that the device is a two-probe configuration, as illustrated in Fig. 1a. In a two-probe configuration the system is divided into a left electrode, central region and a right electrode. For the left and the right electrode we will describe the system using periodic boundary conditions.

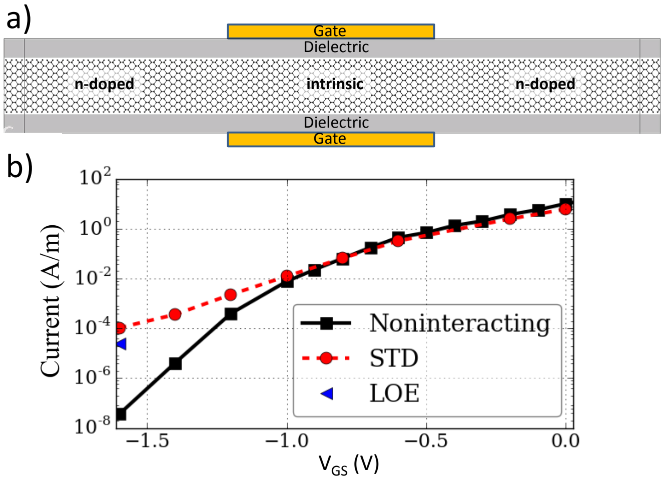

Fig. 1. (a) Silicon n-i-n junction tunneling device with a source and drain doping of $1.0 \times 10^{21}\,\mathrm{cm}^{-3}$, length of 33 nm and $\sim$ 2000 atoms. (b) Current versus gate-voltage, V_G, for a source-drain voltage V_{SD}=0.1 V and at 300 K. The solid line shows results without EPC (Noninteracting), the red dashed line shows results including EPC in the STD approximation, and the single blue triangle shows results for including EPC in the LOE approximation

In the following subsections we will introduce the two methods. Both methods are post-processing methods, i.e. they use the Hamiltonian from a self-consistent electronic structure calculation of the systems without electron-phonon coupling. The first method use a Special Thermal Displacement (STD) of all atoms at a specific temperature which mimics the thermal fluctuations of the configuration. The inelastic current can be obtained by calculating a single elastic transmission spectrum for this STD configuration[4]. The second method is based on lowest order perturbation theory in the electron-phonon coupling self-energies using the Lowest Order Expansion (LOE)[6].

A. The STD method

The STD method is conceptually related to the Special Quasi-Random Structure (SQS) method for describing Random Alloys. In the SQS method a single geometry is used to represent a random alloy, the geometry is selected to give the same physical property as the average value obtained for a random distribution of structures[7]. In the STD method a single configuration is selected which have the same physical property as the average value obtained for a thermal distribution of structures[4], [8]. In Ref. [9] we showed that the average transmission from a thermal distribution of configurations accurately describes the inelastic electron transmission spectrum due to electron-phonon scattering at this temperature. In the STD method the average is replaced with a single representative geometry. To calculate the STD we need the Dynamical matrix (D) of the central region of the system. A first-principles calculation of the full Dynamical Matrix can be very time consuming, however, for most systems a good approximation can be obtained by using a force-field, or in the case of a repeated two-probe system by repeating D calculated for the left electrode.

To obtain the STD we first need to calculate the phonon eigenspectrum using the Dynamical matrix of the central region. The phonon modes are labeled by λ with frequency ω_λ, eigenmode vector $\mathbf{e}_\lambda$, and characteristic length, l_λ. The STD is then given by[4]

$$\mathbf{u}_{STD}(T) = \sum_\lambda s_\lambda (-1)^{\lambda-1} \sigma_\lambda(T) \mathbf{e}_\lambda \qquad (1)$$

Here s_λ denotes the sign of the first non-zero element in $\mathbf{e}_\lambda$ enforcing the same choice of "gauge" for the modes. The Gaussian width σ is related to the mean square displacement $\langle \mathbf{u}_\lambda^2 \rangle = l_\lambda^2 (2 n_B(T) + 1) = \sigma_\lambda^2(T)$ at a temperature T, where n_B is the Bose-Einstein distribution.

The 'trick' in the STD method is the use of opposite phases for phonons with similar frequencies, in this way phonon-phonon correlation functions average to zero and the transmission spectrum of the STD configuration becomes similar to a thermal average of single phonon excitations.

The final step in the STD method is to calculate the self-consistent Hamiltonian of the displaced system, and use that to calculate a normal electron transmission spectrum. Thus, the computation effort for the inelastic transmission is for the STD method similar to elastic transmission.

B. The LOE method

The LOE method for the inelastic current calculation is based on perturbation theory in the first Born approximation. Besides the Hamiltonian and the Dynamical matrix of the central region (as needed in the STD method) it requires knowledge of the Hamiltonian derivative with respect to the atomic positions in the central region, $\nabla H(r)$.

The first-principles calculations of $\nabla H(r)$ can be prohibitive for a large system. The system in Fig. 1a is what we will call a "Repeated Two-probe system". In such a system the atomic configuration of the central region can be generated by repeating the left electrode along the transport direction. For a repeated two-probe system $\nabla H(r)$ can be obtained to a good approximation from the $\nabla H(r)$ of the left electrode.

From t$\nabla H(r)$ of the central region we can get the electron-phonon matrix element in reciprocal space[10].

$$\begin{aligned} M^{\mu\nu}_{\lambda,\mathbf{k},\mathbf{q}} &= \sum_{mn} e^{i\mathbf{k}\cdot(\mathbf{R}_n - \mathbf{R}_m) - i\mathbf{q}\cdot\mathbf{R}_m} \\ &\quad \times \langle \phi_\nu \mathbf{R}_m | \mathbf{v}_{\mathbf{q},\lambda} \cdot \nabla H_0(r) | \phi_\mu \, \mathbf{R}_n \rangle, \end{aligned} \qquad (2)$$

where the (mn)-sum runs over repeated unit cells in the super cell calculation of the hamiltonian derivatives[10], and the subscript 0 indicates that the derivatives are only calculated for atoms in the unit cell with index 0. $|\phi_\mu \, \mathbf{R}_n\rangle$ represent the μ'th LCAO basis orbital in the unit cell displaced from the reference cell by the lattice vector $\mathbf{R}_n$.

Following ref. [6] we obtain inelastic transmission functions, which are symmetric in the applied bias and account for a finite transfer of momentum, $\mathcal{T}^{ems/abs}_{\lambda,\mathbf{k},\mathbf{q}}(\epsilon) =$

$$\begin{aligned} &\mathrm{Tr}\left[M^\lambda \tilde{A}_L^{\mathbf{k}+\mathbf{q}}(\pm) M^\lambda A_R^{\mathbf{k}}(\mp) \right] \\ +\mathrm{Im}\Big\{ &\mathrm{Tr}\left[M^\lambda A_R^{\mathbf{k}+\mathbf{q}}(\pm) \Gamma_L^{\mathbf{k}+\mathbf{q}}(\pm) G^{\mathbf{k}+\mathbf{q}}(\pm) M^\lambda A_R^{\mathbf{k}}(\mp) \right] \\ &+ \mathrm{Tr}\left[M^\lambda A_R^{\mathbf{k}+\mathbf{q}}(\mp) \Gamma_L^{\mathbf{k}+\mathbf{q}}(\mp) G^{\mathbf{k}+\mathbf{q}}(\mp) M^\lambda A_L^{\mathbf{k}}(\pm) \right] \Big\} \end{aligned} \qquad (3)$$

where we use the short hand notation $A_L^{\mathbf{k}}\left(\epsilon \pm \frac{\hbar\omega_\lambda}{2}\right) = A_L^{\mathbf{k}}(\pm)$. In the first line we have $\tilde{A}_L^{\mathbf{k}+\mathbf{q}} = \left(G^{\mathbf{k}+\mathbf{q}}\right)^\dagger \Gamma_L^{\mathbf{k}+\mathbf{q}} G^{\mathbf{k}+\mathbf{q}}$, whereas the other spectral functions are defined as $A_L^{\mathbf{k}+\mathbf{q}} = G^{\mathbf{k}+\mathbf{q}} \Gamma_L^{\mathbf{k}+\mathbf{q}} \left(G^{\mathbf{k}+\mathbf{q}}\right)^\dagger$. We finally evaluate the current following [2] as

$$\begin{aligned} I(V) = -\frac{2e}{hN_{\mathbf{k}}N_{\mathbf{q}}} \sum_{\lambda,\mathbf{k},\mathbf{q}} &\int_{-\infty}^{\infty} d\epsilon [\mathcal{T}^{ems}_{\lambda,\mathbf{k},\mathbf{q}}(\epsilon) F^{ems}_{\lambda,\mathbf{q}}(\epsilon) \\ &+ \mathcal{T}^{abs}_{\lambda,\mathbf{k},\mathbf{q}}(\epsilon) F^{abs}_{\lambda,\mathbf{q}}(\epsilon)], \end{aligned} \qquad (4)$$

where $N_{\mathbf{k}}$ and $N_{\mathbf{q}}$ are the numbers of $\mathbf{k}$ and $\mathbf{q}$-points and where the energy- and mode depended prefactors are

$$\begin{aligned} F^{ems}_{\lambda,\mathbf{q}}(\epsilon) &= n_F(\epsilon - \mu_L)[1 - n_F(\epsilon - \mu_R + \hbar\omega_{\lambda,\mathbf{q}})][n_B(\hbar\omega_{\lambda,\mathbf{q}}) + 1] \\ &\quad - n_F(\epsilon - \mu_R - \hbar\omega_{\lambda,\mathbf{q}})[1 - n_F(\epsilon - \mu_L)] n_B(\hbar\omega_{\lambda,\mathbf{q}}) \\ F^{abs}_{\lambda,\mathbf{q}}(\epsilon) &= n_F(\epsilon - \mu_L)[1 - n_F(\epsilon - \mu_R - \hbar\omega_{\lambda,\mathbf{q}})] n_B(\hbar\omega_{\lambda,\mathbf{q}}) \\ &\quad - n_F(\epsilon - \mu_R + \hbar\omega_{\lambda,\mathbf{q}})[1 - n_F(\epsilon - \mu_L)][n_B(\hbar\omega_{\lambda,\mathbf{q}}) + 1], \end{aligned}$$

where $n_F(\epsilon) = 1/(e^{\epsilon/k_BT} + 1)$ and $n_B(\hbar\omega) = 1/(e^{\hbar\omega/k_BT} - 1)$ are the Fermi-Dirac and Bose-Einstein distribution functions respectively.

In calculating the inelastic current we formally have a sum over all the phonon modes. In order to reduce the computational burden, we perform a summation of the phonon modes in energy intervals to form new effective phonon modes. We typically use intervals of 10 meV length, i.e $[0, 0.01]$ eV, $[0.01, 0.02]$eV, etc. Similar approximations are commonly used in other codes[11]. The sum over modes in Eq. (4) and related equations above is thus replaced by a sum over phonon energy intervals using the redefined phonon modes. Otherwise the formulas remain the same.

III. Results

In the following we compare the methods for the inelastic transport in a nanoscale 2-d silicon n-i-n device and the phonon limited mobility of a silicon nanowire.

A. n-i-n junction

The first system we investigate is a 2-d Silicon n-i-n junction, as illustrated in Fig. 1a. The system is confined in the vertical direction with a thickness of 2 nm and the surfaces are passivated with Hydrogen atoms. The out-of-plane direction is periodic, and the horizontal direction is the transport direction. The device is surrounded by a 1 nm dielectrics with $\epsilon = 4$ and controlled through a gate electrode. The device is intrinsic between the gate electrodes, and n-doped (1.0×10^{21} cm^{-3}) outside the gate.

For the DFT-NEGF calculation we use the ATK package[5]. We use the Local Density Approximation (LDA) for the exchange-correlation functional. The Si and H atoms are described by normconserving pseudopotentials and the electronic structure described by a spd basis set (9 orbital per atom). To describe the periodic direction we use 9 k-points for the SCF calculation and 51 k-points for the transport calculation. The dielectric and gate electrode are described at the continuum level within the poisson equation for the Hartree potential. The method for doping is described in Ref. [12]. The elastic transport at an electronic temperature of 300K is shown with the black solid line in Fig. 1b.

The next step is to include EPC using the STD approximation. To this end we need the Dynamical matrix which we calculate numerically from a classical potential[13]. Using the dynamical matrix we calculate the STD at 300K and perform a SCF calculation for the resulting geometry. From the SCF Hamiltonian we calculate the transmission and the resulting current for different gate biases. The result is shown with the dashed red line in Fig. 1b. We see that the EPC has only little effect for the on-current, however, the off-current is increased by 4 orders of magnitude and the subthreshold swing is degraded from 97 mV/dec to 375 mV/dec.

To check if this result is correct we perform a LOE calculation for the off-current. For this calculation we need the EPC matrix element. Since the system is a repeated two probe device, we only need to calculate the EPC for the left electrode and repeat it to get the EPC of the central region. We can now calculate the inelastic current in the LOE approximation. The LOE calculation is more time consuming and we have only calculated a single point for the off-current. As seen from Fig. 1b the LOE and STD approximations are in excellent agreement. Since the two approaches are very different, it shows that the different approximations for including EPC are valid.

Existing device simulations on silicon FETs have not reported any significant phonon-assisted tunneling. We believe that this is most likely because they either neglect quantum-tunneling, or are based on deformations potentials (corresponding to a purely imaginary and diagonal self-energy in the NEGF formalism) and effective-mass or tight-binding approximations[14], [15], [16], [17]. Thus, our new first principles calculations with a complete description of the EPC are the most accurate description to date of a nanoscale FET. The main error source in the calculation is the use of DFT-LDA which severely underestimates the band gap of Si (0.66 eV in our calculation). We are presently investigating the use of LDA+1/2[18] which reproduces the bandgap of Si (1.18 eV). Preliminary results show that an increased bandgap slightly reduce the phonon assisted tunneling but the main conclusions are the same.

B. Silicon nanowire

In a recent paper[9], we calculated the phonon limited mobility of nanowires using a full first-principles calculations combined with the Boltzmann Transport Equations (BTE)[10] and also from a Molecular Dynamics (MD) Landauer approach. In the MD-Landauer approach an ensemble of configurations are obtained from MD trajectories at the given temperature, and then the transmission spectrum is averaged over this ensemble to obtain an inelastic transmission spectrum[10]. The mobility is obtained through the length dependence of the conductance, i.e. transmission calculations for typically 3 different system lengths are used to calculate the mobility. The STD method can also be considered as a “one-shot” MD-Landauer approach, instead of using an ensemble of configurations, we use a single configuration with the same average atom-atom correlations as the MD ensemble. In the following we will compare the BTE and MD methods with the LOE and STD methods.

The system we consider is a silicon nanowire, illustrated in Fig. 2a. It has a diameter of 1.5nm and is oriented along the (100) direction. The surfaces are passivated with hydrogen. For the MD, STD and LOE methods we consider 3 different lenghts, 5, 10 and 15 nm. Other computational details are similar to the n-i-n study.

Fig. 2b shows the resulting mobilities. The BTE result serves as a good point of reference since it is the most conventional approach. We see that the STD results closely follow the BTE data, proving the validity of the approach. The MD results shows a higher mobility than the STD results and the discrepancy is increased for decreasing temperature. We expect that this is due to the use of classical MD which neglect zero-point motion. The zero-point motion is included in both the BTE and STD calculations. The LOE result gives slightly to high values at low temperatures. For a one-dimensional system the transmission near the band edge is more sensitive towards perturbations due to the van Hove singularity in the density of states. In this limit, one may observe strong modifications of the current beyond lowest order perturbation theory. Recently, several papers have developed an analytic continuation approach that enables a renormalization of the LOE result so that it gives results equivalent to the self-

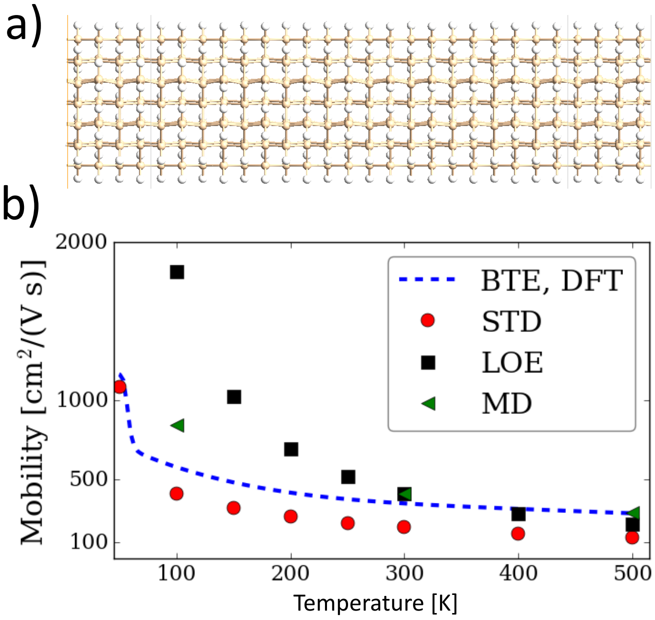

Fig. 2. a) Geometry of the silicon nanowire used for the mobility calculation. b) Mobility as function of temperature for the BTE, STD, MD and LOE methods.

consistent Born approximation[14]. Here we employ the lowest order analytic continuation approach for the nanowire due to its simplicity. While the LOE includes EPC as a perturbation to the current, MD and STD directly evaluates the inelastic current at displaced atomic positions. In general, all four methods gives results that agree from room temperature and above.

IV. Conclusion

We have presented two methods which enable computational efficient first principles transport calculations including EPC. Most efficient and promising is the STD method. This method uses an additional elastic transport calculation for a displaced configuration to provide the effect of EPC, and thereby allows for including EPC at essentially the same cost as elastic quantum transport calculations. We presented results for the electrical properties of a silicon 2D-FET and mobility of a silicon nanowire. The results illustrated the accuracy of the approach.

Acknowledgment

The authors acknowledge support from Innovation Fund Denmark through Grant No. 79-2013-1 and the Quantum Innovation Center (QUBIZ). CNG is sponsored by the Danish National Research Foundation, project No. DNRF103.

References

[1] M. Brandbyge, J.-L. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, "Density-functional method for nonequilibrium electron transport," *Phys. Rev. B*, vol. 65, no. 16, p. 165401, mar 2002.

[2] W. Vandenberghe, B. Sorée, W. Magnus, and M. V. Fischetti, "Generalized phonon-assisted zener tunneling in indirect semiconductors with non-uniform electric fields: A rigorous approach," *Journal of Applied Physics*, vol. 109, no. 12, p. 124503, 2011.

[3] T. Frederiksen, M. Paulsson, M. Brandbyge, and A.-P. Jauho, "Inelastic transport theory from first principles: Methodology and application to nanoscale devices," *Physical Review B*, vol. 75, no. 20, p. 205413, 2007.

[4] T. Gunst, T. Markussen, M. Palsgaard, K. Stokbro, and M. Brandbyge, *Submitted*, 2017.

[5] "Atomistix ToolKit version 2017.3," quantumWise A/S (www.quantumwise.com).

[6] J.-T. Lü, R. B. Christensen, G. Foti, T. Frederiksen, T. Gunst, and M. Brandbyge, "Efficient calculation of inelastic vibration signals in electron transport: Beyond the wide-band approximation," *Physical Review B*, vol. 89, no. 8, p. 081405, 2014.

[7] A. Zunger, S.-H. Wei, L. Ferreira, and J. E. Bernard, "Special quasirandom structures," *Physical Review Letters*, vol. 65, no. 3, p. 353, 1990.

[8] M. Zacharias and F. Giustino, "One-shot calculation of temperature-dependent optical spectra and phonon-induced band-gap renormalization," *Phys. Rev. B*, vol. 94, no. 7, p. 075125, Aug. 2016. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevB.94.075125

[9] T. Markussen, M. Palsgaard, D. Stradi, T. Gunst, M. Brandbyge, and K. Stokbro, "Electron-phonon scattering from green's function transport combined with molecular dynamics: Applications to mobility predictions," *Phys. Rev. B*, vol. 95, p. 245210, 2017.

[10] T. Gunst, T. Markussen, K. Stokbro, and M. Brandbyge, "First-principles method for electron-phonon coupling and electron mobility: Applications to two-dimensional materials," *Physical Review B*, vol. 93, no. 3, p. 035414, 2016.

[11] M. Luisier and G. Klimeck, "Atomistic full-band simulations of silicon nanowire transistors: Effects of electron-phonon scattering," *Physical Review B*, vol. 80, no. 15, p. 155430, 2009.

[12] D. Stradi, U. Martinez, A. Blom, M. Brandbyge, and K. Stokbro, *Phys. Rev. B*, vol. 93, p. 155302, 2016.

[13] J. Tersoff, "Empirical interatomic potential for silicon with improved elastic properties," *Physical Review B*, vol. 38, no. 14, p. 9902, 1988.

[14] N. Cavassilas, M. Bescond, H. Mera, and M. Lannoo, "One-shot current conserving quantum transport modeling of phonon scattering in n-type double-gate field-effect-transistors," *Applied Physics Letters*, vol. 102, no. 1, p. 013508, Jan. 2013. [Online]. Available: http://scitation.aip.org.globalproxy.cvt.dk/content/aip/journal/apl/102/1/10.1063/1.47753

[15] N. Mori, H. Takeda, and H. Minari, "Effects of phonon scattering on electron transport in double-gate MOSFETs," *J Comput Electron*, vol. 7, no. 3, pp. 268–271, Sep. 2008. [Online]. Available: https://link-springer-com.proxy.findit.dtu.dk/article/10.1007/s10825-008-0199-1

[16] A. Svizhenko and M. P. Anantram, "Role of scattering in nanotransistors," *IEEE Transactions on Electron Devices*, vol. 50, no. 6, pp. 1459–1466, Jun. 2003.

[17] S. O. Koswatta, S. J. Koester, and W. Haensch, "On the Possibility of Obtaining MOSFET-Like Performance and Sub-60-mV/dec Swing in 1-D Broken-Gap Tunnel Transistors," *IEEE Transactions on Electron Devices*, vol. 57, no. 12, pp. 3222–3230, Dec. 2010.

[18] L. G. Ferreira, M. Marques, and L. K. Teles, "Approximation to density functional theory for the calculation of band gaps of semiconductors," *Physical Review B*, vol. 78, no. 12, p. 125116, 2008.