First-principles theory of nonlocal screening in graphene

M. van Schilfgaarde and M. I. Katsnelson

Phys. Rev. B **83**, 081409 — Published 17 February 2011

DOI: 10.1103/PhysRevB.83.081409

First-principles Theory of Nonlocal Screening in Graphene

M. van Schilfgaarde[1] and M. I. Katsnelson[2]

[1]*School of Materials, Arizona State University, Tempe AZ*
[2]*Radboud University Nijmegen, Institute for Molecules and Materials, 6525AJ Nijmegen, The Netherlands*

Using the quasiparticle self-consistent GW (QSGW) and local-density (LD) approximations, we calculate the q-dependent static dielectric function of undoped graphene, and derive an effective 2D dielectric function corresponding to screening of point charges. In the $q\to 0$ limit, the 2D dielectric constant is found to scale approximately as the square root of the macroscopic dielectric function. Its value is $\simeq 4$, in agreement with the predictions of Dirac model. At the same time, in contrast with the Dirac model, the dielectric function is strongly dependent on q. The QSGW approximation is shown to describe QP levels very well, with small systematic errors analogous to bulk *sp* semiconductors. Local-field effects are rather more important in graphene than in bulk semiconductors.

PACS numbers: 73.22.Pr, 71.27.+a, 73.22.-f

Graphene is a first truly two-dimensional (2D) crystal, with unique electronic and structural properties (for review, see Refs. 1–5). Screening of electron-electron and electron-impurity interactions in graphene is an important theoretical issue crucial for both many-body effects in electronic structure[6] and for transport properties, especially, for electron scattering by charge impurities[5,7]. An answer to even the basic issue about the ground state of freely suspended graphene is crucially dependent on the value of screened interaction constant since it lies, probably, quite close to the point of exciton instability, but it is still unknown at which side of it[8,9]. A critical value of the impurity charge for the relativistic collapse[10,11], one of the most interesting quantum field effects potentially observable in graphene, is also dependent on the screened coupling constant. Numerous works[7,10,12–19] treat this issue within the two-band Dirac model. But the Dirac model does not take into account the many other bands involved, which can include van Hove singularities in electron density of states[3] that may possibly be very essential, specifically for screening[20]. Here we develop a definition for an effective 2D dielectric function in an *ab initio* context, and calculate it within the quasiparticle self-consistent GW (QSGW) and local-density (LD) approximations. The former takes into account many-body effects beyond the density functional GGA or LDA schemes essential for correct description of excited states and thus screening effects[21,22].

There are several GW calculations for graphene[23–25], where G and the screened Coulomb interaction W are computed from the LDA. They all predict a notable (20-40%) increase of the Fermi velocity v_F at the Dirac point K relative to the LDA(GGA) value, with v_F between 1.1 and $1.2{\cdot}10^6$ m/s, in very good agreement with experiment[1–4]. The dielectric function and optical conductivity as a function of frequency ω for zero wave vector $\mathbf{q}$=0 was also calculated in Refs. 23,24. Here we focus on the static dielectric function (ω=0) as a function of q. As mentioned above, this quantity is relevant for calculations of resistivity via charge impurities[5,7], as well as for the problem of supercritical Coulomb centers[10,11] and possible exciton instabilities[8,9].

The inverse dielectric function $\epsilon^{-1}(\mathbf{r},\mathbf{r}',\omega)$ relates the change in total potential δV to an external perturbing potential $\delta V^{\rm ext}$ as[21,22]

$$\delta V(\mathbf{r},\omega) = \int d\mathbf{r}'\epsilon^{-1}(\mathbf{r},\mathbf{r}',\omega)\delta V^{\rm ext}(\mathbf{r}',\omega). \qquad (1)$$

ϵ^{-1} is obtained from a convolution of the polarization operator Π and the bare Coulomb interaction v as

$$\epsilon^{-1} = (1 - v\Pi)^{-1}.$$

In a system with translation symmetry, ϵ^{-1}, Π, and v can be expanded in Bloch functions $\{B_I^{\mathbf{q}}(\mathbf{r})\}$, e.g.

$$\epsilon^{-1}(\mathbf{r},\mathbf{r}',\omega) = \sum\nolimits_{\mathbf{q}IJ} B_I^{\mathbf{q}}(\mathbf{r})\epsilon^{-1}_{IJ}(\mathbf{q},\omega)B_J^{\mathbf{q}*}(\mathbf{r}') \qquad (2)$$

The most common choice of $\{B_I^{\mathbf{q}}(\mathbf{r})\}$ are plane waves,

$$B_I^{\mathbf{q}}(\mathbf{r}) \to B_{\mathbf{G}}^{\mathbf{q}}(\mathbf{r}) = \exp(i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}), \qquad (3)$$

$\mathbf{G}$ being reciprocal lattice vectors.

Quantities of interest are coarse-grained averages of $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)$. The "macroscopic" response to a plane wave perturbation is[22]

$$\epsilon_M(\mathbf{q},\omega) = \left[\epsilon^{-1}_{\mathbf{G}=\mathbf{0},\mathbf{G}'=\mathbf{0}}(\mathbf{q},\omega)\right]^{-1} \qquad (4)$$

The matrix structure of ϵ^{-1} with $\mathbf{G}\neq\mathbf{G}'$ reflects local field effects in terms of classical electrodynamics. The quantity $\epsilon_M(\mathbf{q})$ is commonly approximated by just $\epsilon(\mathbf{q})$; that is, the Umklapp processes, or local field effects are neglected. This is not such a bad approximation in *sp* semiconductors but as we show here, it is a rather poor approximation in graphene. $\epsilon_M(\mathbf{q})$ corresponds to screening potential $\delta V^{\rm ext}$ with a single Fourier component $\mathbf{q}$. Selecting $\mathbf{G}$=$\mathbf{G}'$=0 averages ϵ^{-1} over the unit cell, restricting the spatial variation to the envelope $\exp(i\mathbf{q}\cdot\mathbf{r})$. While ϵ_M is a quantity of relevance to some experiments, perhaps the most relevant is screening of a

point charge in the graphene sheet, which governs e.g., scattering from impurities.

As graphene is a 2D system, we need to consider how the impurity potential $v(q) = 4\pi/q^2$ is screened in the sheet. The (statically) screened potential from a point charge at the origin may be written in cylindrical coordinates $\mathbf{r}=(\rho, z, \theta)$ and $\mathbf{q}=(\bar{q}, q_z, \theta_q)$ as

$$W(\rho, z) = \frac{1}{2\pi}\int_0^\infty d\bar{q}\,\bar{q} J_0(\bar{q}\rho)\, W^{2\mathrm{D}}(\bar{q}, z) \quad (5)$$

$$W^{2\mathrm{D}}(\bar{q}, z) = 4\int_0^\infty dq_z e^{iq_z z}\frac{\epsilon^{-1}(\bar{q}, q_z)}{q_z^2+\bar{q}^2} \quad (6)$$

Graphene is hexagonal, and ϵ^{-1} does not depend on θ_q for small $\bar{q}$. $W^{2\mathrm{D}}(\bar{q}, z)$ is the 2D (Hankel) transform of $W(\mathbf{r})$, the analog of the 3D transform $W(\mathbf{q})=\epsilon^{-1}(\mathbf{q})v(\mathbf{q})$. In the absence of screening $\epsilon^{-1}=1$ and $W^{2\mathrm{D}}(\bar{q}, z)$ reduces to the bare coulomb interaction $v^{2\mathrm{D}}(\bar{q}, z)$:

$$v^{2\mathrm{D}}(\bar{q}, z) = 4\int_0^\infty dq_z e^{iq_z z}\frac{1}{q_z^2+\bar{q}^2} = \frac{2\pi}{\bar{q}}e^{-\bar{q}z} \quad (7)$$

An appropriate definition of an effective 2D dielectric function is then

$$\epsilon^{2\mathrm{D}}(\bar{q}, z) = v^{2\mathrm{D}}(\bar{q}, z)/W^{2\mathrm{D}}(\bar{q}, z) \quad (8)$$

This definition corresponds to a typical physical problem for graphene when both interacting charges lie at the same plane (e.g., electrons and holes in the problem of exciton instability). Also, this is a quantity which can be compared with predictions of the Dirac model. Graphene wave functions have some extent in z which must be integrated over to obtain a scattering matrix element. But the largest contribution originates from z=0, so $W^{2\mathrm{D}}(\bar{q}, 0)$ is a reasonable estimate for the scattering potential. This is particularly so for small $\bar{q}$ of primary interest here.

In practice we carry calculations in a periodic array of graphene sheets in the xy plane, spaced by a distance large enough that the sheets interact negligibly. To calculate $\epsilon^{-1}_{\mathbf{G}=\mathbf{G}'=0}(\mathbf{q}, \omega)$ we adopt the all-electron, augmented wave implementation that was developed for the quasiparticle self-consistent GW (QSGW) approximation, described in Ref.[26]. It makes no pseudo- or shape- approximation to the potential, and does not use PWs (Eq. 3) for the product basis $\{B\}$, but a mixed basis consisting of products of augmented functions in augmentation spheres, and plane waves in the interstitial region. The all-electron implementation enables us to properly treat core states. We calculate $\epsilon^{-1}(\mathbf{q}, \omega)$ in the random phase approximation, using Bloch functions for eigenstates[21]. These are obtained from single-particle eigenfunctions $\Psi_{\mathbf{k}n}$ and eigenvalues $\epsilon_{\mathbf{k}n}$ in both the LDA and QSGW approximations. In both cases the generalized LMTO method is used[27,28].

QSGW has been shown to be an excellent predictor of materials properties for many classes of compounds composed of elements throughout the Periodic Table, with unprecedented ability to consistently and reliably predict materials properties over a wide range of materials[26,28]. Nevertheless there are small, systematic errors: in particular bandgaps in insulators such as GaAs, $SrTiO_3$ and NiO, are systematically overestimated. Its origin can be traced to a large extent to the RPA approximation to the polarizability, $\Pi^{\mathrm{RPA}}=iGG$. The RPA bubble diagrams omit electron-hole interactions in their intermediate states. Short-range attractive (electron-hole) interactions induce redshifts in $\mathrm{Im}\,\epsilon(\mathbf{q}, \omega)$ at energies well above the fundamental bandgap; see e.g. Fig. 6 in Ref.[26]. That ladder diagrams are sufficient to remedy most of the important errors in Π^{RPA} was demonstrated rather convincingly in Cu_2O, by Bruneval et al.[29]. Moreover Shishkin et al[30] incorporated these ladder diagrams in an approximate way for several *sp* semiconductors, and established that they do in fact largely ameliorate the gap errors. Yang et al. investigated the effect of ladder diagrams in graphene and graphite, and showed that in a manner very analogous to ordinary semiconductors, these diagrams induce a redshift in the peak of $\mathrm{Im}\,\epsilon^{\mathrm{RPA}}(\mathbf{q}, \omega)$ near 5 eV,[24] of ∼0.6 eV. They found a strong similarity with conventional semiconductors, namely that the redshift from ladder diagrams approximately cancels the error in the LDA joint density of states.

TABLE I: Energy gap E_G and valence bandwidth Γ_{1v} in diamond (eV); Fermi velocities v_F in graphite and graphene (10^6m/sec). There is a significant renormalization of the bandgap from the electron-phonon interaction in diamond, estimated to be 370 meV[31]. Thus QSGW overestimates E_G by a slightly smaller amount than in other semiconductors, and the scaling of Σ as described in the text results in a slightly underestimated gap. The electron-phonon interaction also reduces the Fermi velocity in graphene, estimated to be 4 to 8% in an LDA-linear response calculation[32]. The calculated Fermi velocities should be reduced by this much when comparing to experiment. v_F calculated by QSGW is slightly overestimated, for much the same reason semiconductor gaps are overestimated. v_F calculated from the scaled-Σ potential, is slightly larger than v_F calculated LDA-based GW, i.e. $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$[23], just as semiconductor bandgaps are slightly larger. When renormalized by the electron-phonon interaction, v_F agrees very well with the measured value[33].

	LDA	QSGW	scaled Σ	Expt
E_G, diamond	4.09	5.93	5.56	5.50
Γ_{1v}, diamond	21.3	23.1	22.7	23.0 ± 0.2^a
Γ_{1v}, graphene	19.4	22.9	22.2	
v_F(H), graphite	0.77	0.99	0.94	0.91 ± 0.15
v_F(K), graphene	0.82	1.29	1.20	1.1

[a] Ref. 34

A redshift in the peak of $\mathrm{Im}\epsilon(\omega)$ increases the static dielectric constant ϵ_∞, as can be readily seen through the Kramers-Kronig relations. Remarkably, ϵ_∞ as calculated by the RPA in QSGW, is underestimated by a

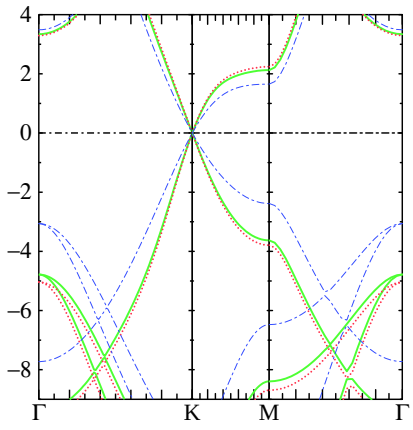

FIG. 1: QSGW bands of graphene (dotted red lines), compared to LDA results (dashed blue lines) and QSGW results with Σ scaled by 0.8 (solid green lines) described in the text. The linear dispersion near K (or H, in graphite) is significantly larger in the QSGW case. Differences are quantified in Table I. The lowest lying unbound state can be seen as a parabolic band starting at Γ near 3.5 eV. It corresponds to the work function. LDA and QSGW work functions are very similar, consistent with the observation that LDA predicts work functions rather well in many systems.

nearly *universal factor* of 0.8, for many kinds of insulators and semiconductors[28], including transition metal oxides such as NiO[26], CeO_2, and *sp* semiconductors[30]. (This error is often approximately canceled in the LDA, fortuitously. As Yang et al. noted, the cancellation seems to apply to graphene in a manner similar to ordinary semiconductors.) Because ϵ is systematically *underestimated*, $W = \epsilon^{-1}v$ and $\Sigma = -iGW$ are systematically *overestimated*; therefore QP excitation energies are also systematically overestimated. We have found that simply scaling by 0.8 (the nearly universal ratio $\epsilon_\infty^{QSGW}/\epsilon_\infty^{expt}$) largely eliminates discrepancies between QSGW and measured QP levels in a wide range of *spd* systems, including all zincblende semiconductors, and many other kinds of insulators. For graphene, we find that the QSGW macroscopic ($q\to 0$) dielectric constant was found to be 80% of the LDA one, consistent with the universal pattern in bulk insulators noted above. The many points of consistency with 3D behavior, both in the QSGW QP levels and the dielectric response suggest that QSGW will exhibit the same reliable description of the 2D graphene system, with similar systematic errors. To confirm this, some band parameters for three pure (undoped) carbon compounds calculated by QSGW and QSGW with Σ scaled by 0.8 are shown in Table I. Scaling QSGW has a minor effect on the quasiparticle levels: e.g. it reduces v_F by 7%. As Table I shows, v_F falls in very close agreement with experiment when Σ is scaled and the electron-phonon interaction is taken into account, consistent with agreement in gaps in the bulk insulators. Even though the QSGW and LDA work functions are similar (Fig. 1), the valence band is significantly widened relative to LDA,[34] more so in graphene than in diamond.

Careful checks for convergence were made in various parameters. To check supercell artifacts, a "small" 3D unit cell with the graphene planes repeated at a spacing equivalent to 4 atomic layers of graphite (25 a.u.) was compared against a "large" cell, with graphene planes spaced at 8 layers. The bands from $-\infty$ to E_F+5 eV were found to be a very similar, with a slight increase in v_F (1.23→1.29 $\cdot 10^6$ m/s). k convergence in the construction of Σ was monitored by comparing QP levels generated on a 6×6×2 k mesh to a 9×9×2 mesh. QP levels were nearly identical: v_F differed by <1% in the both the small and large 3D cells.

$\epsilon_{00}^{-1}(\mathbf{q},\omega)$ must be integrated with a fine k mesh. To deal with the delicate $\mathbf{q}\to 0$ limit, we calculated ϵ^{-1} integrating on a standard k mesh including Γ, and an offset mesh (Eqns. 47 and 52 in Ref. 26), and averaged them. We present data for averaged 18×18×4 meshes. Calculations without local fields were also performed for a pair of 24×24×4 meshes. $\epsilon(\mathbf{q}_{||}, q_z{=}0, \omega{=}0)$ calculated by 18- and 24- (averaged) mesh integrations were essentially indistinguishable for $q{>}0.1{\times}2\pi/a$, and differed by a few percent for $q{>}0.02{\times}2\pi/a$.

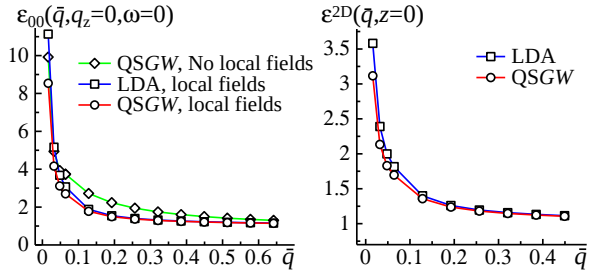

FIG. 2: (Top) Static dielectric function $\epsilon_{00}(\bar{q}, q_z{=}0)$ along the (100) line in graphene, with local fields included and without. $\bar{q}$ is in units of $2\pi/a{=}2.56\text{Å}^{-1}$. The $q\to 0$ limit is delicate and there is some uncertainty in its value. Shown for comparison is the same function calculated from the LDA potential. In the $\bar{q}\to 0$ limit, ϵ_{00} calculated by QSGW is ∼0.8 smaller than the LDA result, similar to the ratio found in bulk semiconductors. (Bottom) Effective layer dielectric function $\epsilon^{2D}(\bar{q}, z{=}0)$ as defined by Eq. (8), with local fields, calculated within the QSGW and LD approximations. Local fields significantly reduce ϵ_{00}. The LDA result for $\epsilon_{00}(\bar{q}{=}0.086, q_z{=}0)$ without local fields is ≃4, which agrees with the $\omega\to 0$ limit in Fig. 11 of Ref.35.

$\epsilon_{00}(\bar{q}, q_z)$ was calculated on a grid of points $\{\bar{q}, q_z\}$; the $q_z{=}0$ case is shown in the first panel of Fig. 2. It was found that ϵ_{00} is well parametrized (max error <0.1) by

$$\epsilon_{00}^{-1}(\bar{q}, q_z) = \frac{a^2(\bar{q}) + q_z^2}{\epsilon_{00}(\bar{q}, 0)\, a^2(\bar{q}) + q_z^2} \qquad (9)$$

$$a^2(\bar{q}) = \frac{a_0 a_1 \bar{q}^2}{a_1 + \bar{q}^2} \approx a_0 \bar{q}^2 \qquad (10)$$

where a_0=1.3 and 1.2 for QSGW and LDA, respectively, and $a_1{=}1.6(2\pi/a)^2$. The approximate form for

a in Eq. 10 is sufficient for any q where ϵ_{00} differs significantly from unity. With Eq. (9) W^{2D} can be integrated analytically. Taking the approximate expression for $a^2(\bar{q})$ we obtain

$$\epsilon^{2D}(\bar{q},z) = \frac{\gamma(\gamma^2-1)}{\gamma(a_0-1)+(\gamma^2-a_0)e^{(1-\gamma)\bar{q}z}} \qquad (11)$$

where $\gamma = \sqrt{a_0\epsilon_{00}(\bar{q},0)}$.

Fig. 2 shows both kinds of dielectric functions, ϵ_M corresponding to the macroscopic polarizability, and the effective 2D static dielectric function $\epsilon^{2D}(\bar{q},z{=}0)$ calculated from Eq. (9). Local fields reduce the strength of the screening. The difference between LDA and QSGW results are modest; and as noted earlier, the LDA results are likely to be slightly better because they benefit from a fortuitous cancellation of errors. As $\bar{q}\to 0$, γ is significantly larger than a_0 and unity. Keeping only the leading order in γ, we obtain the surprising result that $\epsilon^{2D}(0,z{=}0)\approx\sqrt{a_0\epsilon_{00}(\bar{q},q_z{=}0)}$. $\epsilon^{2D}(0,z{=}0)$ is quite close to the prediction of the Dirac model[7,12,14]. Such a model predicts $\epsilon(q) = 1+\pi e^2/2\hbar v_F \approx 4.4$ independent of q. We find $\epsilon^{2D}(\bar{q},z{=}0)\approx 3.5$ for $\bar{q}\to 0$, but, in an essential difference with the Dirac model, ϵ^{2D} is a very strong function of $\bar{q}$.

The case of small $\bar{q}$ ($\bar{q}\sim k_F \leq 10^{-2}$ Å^{-1}) is relevant for transport properties. In this region our first-principles calculations do not dramatically contradict predictions of the Dirac model. At the same time, for the problem of supercritical Coulomb centers and relativistic collapse (fall on the center)[10,11] distances of order of several inverse lattice constants are essential (this is the radius of screening cloud, according to renormalization group analysis[10]), which corresponds to larger q. They are essential as well for the problem of excitonic instabilities[8,9]. For this region our results show that the Dirac model *overestimates* the screening.

Acknowledgments

MIK acknowledges support from Stichting voor Fundamenteel Onderzoek der Materie (FOM), the Netherlands. MvS was supported by ONR contract N00014-7-1-0479 and NSF QMHP-0802216. We thank F. Guinea for helpful discussions.

[1] A. K. Geim and K. S. Novoselov, Nature Mater. **6**, 183 (2007).
[2] M. I. Katsnelson, Mater. Today **10**, 20 (2007).
[3] A. H. C. Neto et al, Rev. Mod. Phys. **81**, 109 (2009).
[4] A. K. Geim, Science **324**, 1530 (2009).
[5] S. Das Sarma, S. Adam, E. H. Hwang, and E. Rossi, preprint arXiv:1003.4731.
[6] J. Gonz'ales, F. Guinea, and M. A. H. Vozmediano, Nucl. Phys. B **424**, 596 (1994).
[7] T. Ando, J. Phys. Soc. Japan **75**, 074716 (2004).
[8] D. V. Khveshchenko, Phys. Rev. Lett. **87**, 246802 (2001).
[9] J. Sabio, F. Sols, and F. Guinea, Phys. Rev. B **81**, 045428 (2010); *ibid*, arXiv:1007.3471.
[10] A. V. Shytov, M. I. Katsnelson, and L. S. Levitov, Phys. Rev. Lett. **99**, 236801 (2007); *ibid*, Phys. Rev. Lett. **99**, 246802 (2007).
[11] V. M. Pereira, J. Nilsson, and A. H. Castro-Neto, Phys. Rev. Lett. **99**, 166802 (2007).
[12] B. Wunsch et al, New J. Phys. **8**, 318 (2006).
[13] M. I. Katsnelson, Phys. Rev. B **74**, 201401(R) (2006).
[14] E. H. Hwang and S. Das Sarma, Phys. Rev. B **75**, 205418 (2007).
[15] M. M. Fogler, D. S. Novikov, and B. I. Shklovskii, Phys. Rev. B **76**, 233402 (2007).
[16] M. Polini, A. Tomadin, R. Asgari, and A. H. MacDonald, Phys. Rev. B **78**, 115426 (2008).
[17] E. Rossi and S. Das Sarma, Phys. Rev. Lett. **101**, 166803 (2008).
[18] L. Brey and H. A. Fertig, Phys. Rev. B **80**, 035406 (2009).
[19] M. M. Fogler, Phys. Rev. Lett. **103**, 236801 (2009).
[20] M. I. Katsnelson and A. V. Trefilov, Phys. Lett. A 109, 109 (1985); *ibid*, Phys. Rev. B 61, 1643 (2000).
[21] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys **61**, 237 (1998).
[22] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).
[23] P. E. Trevisanutto et al, Phys. Rev. Lett. **101**, 226405 (2008).
[24] L. Yang et al, 186802 (2009).
[25] C. Attacalite and A. Rubio, Phys. Stat. Sol. (b) **246**, 2523 (2009).
[26] T. Kotani, M. van Schilfgaarde, and S. V. Faleev, Phys. Rev. B **76**, 165106 (2007).
[27] To ensure completeness in the interstitial floating orbitals were placed in the "empty sites" where nucleii would fall if graphene were turned into graphite. An *spdf spd* basis was used for C atoms.
[28] S. V. Faleev, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. 93, 126406 (2004); M. van Schilfgaarde, T. Kotani, S. V. Faleev, Phys. Rev. Lett. **96**, 226402 (2006); *ibid*, Phys. Rev. B **74**, 245125 (2006).
[29] F. Bruneval et al, Phys. Rev. Lett. **97**, 267601 (2006).
[30] M. Shishkin, M. Marsman, and G. Kresse, Phys. Rev. Lett. **99**, 246403 (2007).
[31] M. Cardona and M. L. W. Thewalt, Rev. Mod. Phys. **77**, 1173 (2005).
[32] C.-H. Park et al, Phys. Rev. Lett. **99**, 086804 (2007).
[33] Y. Zhang et al, Nature **438**, 201 (2005).
[34] I. Jiménez et al, Phys. Rev. B **56**, 7215 (1997).
[35] A. G. Marinopoulos et al, Phys. Rev. B **69**, 245419 (2004).