Effect of spatial nonlocality on the density functional band gap

Myrta Grüning,[1,*] Andrea Marini,[2] and Angel Rubio[3]

[1]*Donostia International Physics Center (DIPC) and European Theoretical Spectroscopy Facility (ETSF), E-20018 San Sebastián, Spain*
[2]*CNR-INFM and European Theoretical Spectroscopy Facility (ETSF), Dipartimento di Fisica dell'Universita di Roma*
"Tor Vergata," Via della Ricerca Scientifica, I-00133 Roma, Italy
[3]*Departamento de Física de Materiales, Facultad de Químicas Universidad del Pais Vasco, European Theoretical Spectroscopy Facility*
(ETSF), Centro Mixto CSIC-UPV, Donostia International Physics Center (DIPC), E-20018 Donostia-San Sebastián, Spain

We show that for a large class of exchange-correlation functionals the local exchange-correlation potential obtained within an optimized effective potential severely underestimates the band gap. On the other hand, the corresponding nonlocal potential obtained from a generalized Kohn-Sham scheme provides a much better description of the band gap, in good agreement with experiments. These results strongly indicate that a local exchange-correlation potential, however good the exchange-correlation approximation, cannot capture the delicate interplay between correlation effects and spatial localization in the KS band structure, unless the (cumbersome) contribution from the derivative discontinuity of the exchange-correlation energy functional is considered.

Since the fundamental Kohn-Sham (KS)[1] paper much effort has been made to devise a consistent and efficient scheme to deal with the band gap problem in density-functional theory (DFT). That is, the systematic underestimation of the first electronic excitation energy in semiconductors and insulators. It has been demonstrated that in extended systems[2,3] the KS band gap does not describe the fundamental gap because of the continuous dependence of the exchange-correlation (XC) potential upon the electronic density.[4] In fact standard approximations for v_{XC}—local density (LDA) or generalized gradient approximations—underestimate the gap by ~ 30–50% (Ref. 5), and similar results have been obtained with more accurate approximations,[6] based on the many-body perturbation theory (MBPT). Despite these results, the success of the exchange-only optimized effective potential (OEP) method in describing the gap of simple semiconductors[7] poses the question about the existence of *practical* approximations that can successfully predict the band structure of solids. The OEP method[8–10] provides improved approximations for the XC potential from a given level of accuracy of the XC energy functional. The exchange-only scheme corresponds to neglect correlation effects suggesting that they are not important in the DFT description of the fundamental gap. However this is in contrast to the common MBPT wisdom.[11] More importantly, exchange-only OEP does not perform equally well for wide-gap insulators,[12] and including correlation effects shrinks back the gap to the LDA value, as it has been confirmed in the self-consistent OEP calculations at *GW* level of approximation.[13]

A possible explanation of the poor performance of the OEP method could be the approximations used for describing the correlation part of the potential. However, the quest for better descriptions of the correlation functional means, in the many-body language, higher perturbation orders, higher complexity, and, very often, not feasible expressions. Instead of attempting such a costly approach, we consider an alternative density-functional method to devise expressions for the XC potential: the generalized KS (GKS) scheme.[14] Un-

like the OEP method, the GKS scheme treats part of the exchange-correlation effects using a nonlocal potential. The GKS has been used in combination with empirical Hartree-Fock hybrid functional,[15] or simple models for the screened exchange[16,17] providing encouraging results for electronic and structural properties, especially of semiconductors (where the exchange-only OEP performs equally well).

In this work we systematically compare the GKS and OEP-KS results for the fundamental gaps of Si (semiconductor), LiF, and Ar (wide-gap insulators) using a given approximation for the XC energy functional. The results show that the GKS approach yields gaps in much better agreement with the experiment than the OEP-KS scheme, where the XC potential is imposed to be local in space. We trace back the poor performance of the OEP-KS scheme as due to the delicate interplay between correlation effects and spatial localization that is not properly described using a local potential.

We consider the XC energy functional $E_{XC}=E_{XC}^{orb}+E_{XC}^{den}$ of an interacting system composed of N electrons. E_{XC}^{orb} is the orbital-dependent term,[18]

$$E_{XC}^{orb} = -\frac{1}{2}\int d\mathbf{r}d\mathbf{r}' n(\mathbf{r}',\mathbf{r})w(\mathbf{r},\mathbf{r}')n(\mathbf{r},\mathbf{r}'), \qquad (1)$$

with $n(\mathbf{r},\mathbf{r}')=\sum_i^N \psi_i(\mathbf{r})\psi_i^*(\mathbf{r}')$ the one-particle density matrix of the auxiliary system. w describes the part of the electron-electron interaction treated nonlocally while the reminder is included in the explicitly density-dependent term E_{XC}^{den}. The OEP approach derives a KS equation corresponding to the E_{XC} functional

$$\left(-\tfrac{1}{2}\nabla^2 + v_{ext} + v_H + v_{XC}^{den} + v_{XC}^{orb}\right)\psi_i^{KS} = \varepsilon_i^{KS}\psi_i^{KS}. \qquad (2)$$

In Eq. (2), v_{ext} is the external potential, v_H is the Hartree potential, and $v_{XC}^{den}=\delta E_{XC}^{den}/\delta n(\mathbf{r},\mathbf{r})$. The local potential v_{XC}^{orb} corresponding to E_{XC}^{orb} is obtained via the OEP equation,[8–10]

$$\int d\mathbf{r}' d\mathbf{r}'' [v_{\text{XC}}^{\text{NL}}(\mathbf{r}',\mathbf{r}'') - v_{\text{XC}}^{\text{orb}}(\mathbf{r}')\delta(\mathbf{r}'-\mathbf{r}'')]$$

$$\times \sum_{ij} \frac{\psi_i(\mathbf{r})\psi_i^*(\mathbf{r}')\psi_j^*(\mathbf{r})\psi_j(\mathbf{r}'')}{\varepsilon_i - \varepsilon_j} + \text{c.c.} = 0. \quad (3)$$

This integral equation relates the local potential to the non-local potential $v_{\text{XC}}^{\text{NL}}$ defined as

$$v_{\text{XC}}^{\text{NL}}(\mathbf{r},\mathbf{r}') = \frac{\delta E_{\text{XC}}^{\text{orb}}}{\delta n(\mathbf{r},\mathbf{r}')} = - n(\mathbf{r}',\mathbf{r})w(\mathbf{r},\mathbf{r}'). \quad (4)$$

Within the GKS scheme the orbital-dependent part, Eq. (1), of the functional is interpreted as interaction energy among the particles of the auxiliary system and treated with a non-local potential within the set of one-particle equations

$$\left(-\tfrac{1}{2}\nabla^2 + v_{\text{ext}} + v_H + v_{\text{XC}}^{\text{den}}\right)\psi_i^{\text{GKS}} + \int d\mathbf{r}' v_{\text{XC}}^{\text{NL}}(\mathbf{r},\mathbf{r}')\psi_i^{\text{GKS}}(\mathbf{r}')$$

$$= \varepsilon_i^{\text{GKS}}\psi_i^{\text{GKS}}. \quad (5)$$

Equation (5) has the form of a quasiparticle equation, but the wave function of the system is constrained to be a Slater determinant, and part of the XC effects are treated locally. The local XC potential $v_{\text{XC}}^{\text{den}}$ ensures that the density of the GKS system equals the density of the full-interacting system.[14]

Equations (3) and (5) make clear the crucial difference between the OEP-KS and GKS schemes: the $E_{\text{XC}}^{\text{orb}}$ term enters the one-particle equation with a local potential within OEP ($v_{\text{XC}}^{\text{orb}}$), and with a nonlocal potential ($v_{\text{XC}}^{\text{NL}}$) within GKS.

In our calculation we approximate the XC energy with the Coulomb-hole screened-exchange (CHSX+) functional. The CHSX+ approximation is derived from the GW[19,20] level of the MBPT including the effects of the self-energy energy dependence as a renormalization factor.[21] Unlike the hybrid functionals which need an empirical parameter to define the partition into a density and an orbital part, the CHSX+ is naturally partitioned into a density (Coulomb hole) and an orbital-dependent (screened exchange) part. The Coulomb hole has the form

$$E_{\text{XC}}^{\text{CH}} = \frac{1}{2}\int d\mathbf{r} d\mathbf{r}'\, \delta(\mathbf{r}-\mathbf{r}')n(\mathbf{r},\mathbf{r}')\left[w(\mathbf{r},\mathbf{r}') - \frac{1}{|\mathbf{r}-\mathbf{r}'|}\right]. \quad (6)$$

The interaction w in Eqs. (1) and (6) is

$$w(\mathbf{r},\mathbf{r}') = \int d\mathbf{r}'' \frac{\epsilon^{-1}(\mathbf{r},\mathbf{r}'')}{|\mathbf{r}''-\mathbf{r}'|}, \quad (7)$$

where ϵ^{-1} is the inverse of the static screening. We calculate ϵ^{-1} in the random-phase approximation using the wave functions and eigenvalues of the local density approximation (LDA) calculation and we consider it as fixed and independent of the density.[22]

We implemented the CHSX+ functional within the OEP-KS and the GKS schemes into the DFT module of the plane-wave SELF[23] code. In both schemes we used the LDA eigenfunctions as a basis set. The LDA[24] calculations have
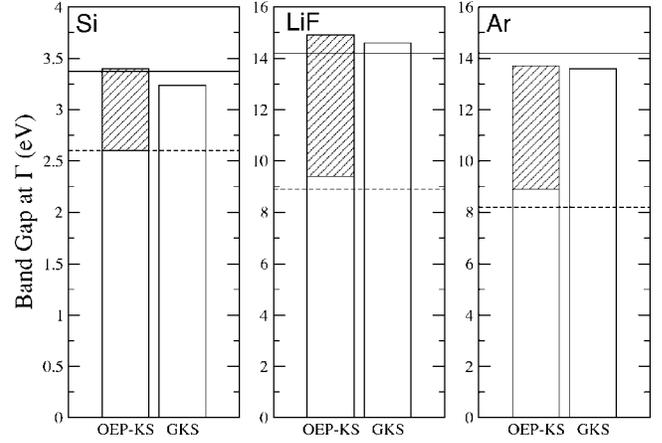

FIG. 1. Direct band gap at Γ for Si, LiF, and Ar evaluated from the OEP-KS and the GKS eigenvalue differences using the CHSX + approximation. The striped bar is $\Delta_{\text{XC}}^{\text{orb}}$ contribution to the gap evaluated at the first order perturbation theory [Eq. (14)]. The continuous (dashed) line indicates the experimental (Refs. 30–32) (LDA) band gap.

been performed with the ABINIT[25] code. We have used Troullier-Martins[26,27] norm conserving LDA pseudopotentials. For bulk Si, LiF, and Ar (face-centered cubic lattice at experimental lattice constant) the plane wave cutoff calculations are 18 Ry, 80 Ry, and 70 Ry, respectively. For all systems the Brillouin zone is sampled with a $4 \times 4 \times 4$ Monkhorst-Pack[28] grid.[29]

Figure 1 summarizes our results. The CHSX+ within the GKS improves substantially on the LDA, yielding an excellent estimate of the experimental band gap[30–32] for the three systems considered: Si (a semiconductor), LiF, and Ar (two wide-gap insulators). On the contrary, the CHSX+ within the OEP-KS scheme yields band gaps almost indistinguishable from the LDA values, strongly underestimated as compared with the experimental gap. More importantly, the OEP and GKS band gaps refer to the same XC energy functional, thus the large discrepancy in the two series of results is entirely due to the different treatment of the spatial locality of the XC potential. These results can be analyzed in more detail by looking at the relation between the definition of the true energy gap, the OEP-KS, and the GKS gap.[33]

Within the KS scheme (in the following discussion KS stands for OEP-KS) the orbital-dependent part contributes to the effective local potential in Eq. (2). The eigenvalue band gap is the fundamental gap of a system of noninteracting particles in such effective potential. Therefore the gap $\varepsilon_{\text{gap}}^{\text{KS}}$ is entirely due to the derivative discontinuity Δ_s of the KS kinetic energy[2,3] T_s

$$\varepsilon_{\text{gap}}^{\text{KS}} = \Delta_s \equiv \lim_{\delta \to 0}\left(\left.\frac{\delta T_s}{\delta n}\right|_{N+\delta} - \left.\frac{\delta T_s}{\delta n}\right|_{N-\delta}\right). \quad (8)$$

Within the GKS scheme the orbital-dependent part enters the one-particle equations [Eq. (5)] as a nonlocal potential and therefore its derivative discontinuity $\Delta_{\text{XC}}^{\text{orb}}$ [defined as Δ_s in Eq. (8)] contributes to the gap

$$\varepsilon_{\text{gap}}^{\text{GKS}} = \Delta_s + \Delta_{\text{XC}}^{\text{orb}}. \tag{9}$$

Finally, both within the KS and within the GKS scheme the total energy of the full-interacting system has the same functional form

$$E = T_s + \int d\mathbf{r}\, n(\mathbf{r})[v_{\text{ext}}(\mathbf{r}) + v_H(\mathbf{r})] + E_{\text{XC}}^{\text{den}} + E_{\text{XC}}^{\text{orb}}, \tag{10}$$

with T_s and $E_{\text{XC}}^{\text{orb}}$ evaluated starting from either the KS or the GKS wave functions. From Eq. (10) it follows that the true gap E_{gap} is

$$E_{\text{gap}} = \Delta_s + \Delta_{\text{XC}}^{\text{orb}} + \Delta_{\text{XC}}^{\text{den}}. \tag{11}$$

Comparing Eq. (11) with Eqs. (8) and (9) it follows that

$$E_{\text{gap}} = \varepsilon_{\text{gap}}^{\text{KS}} + \Delta_{\text{XC}}^{\text{den}} + \Delta_{\text{XC}}^{\text{orb}} = \varepsilon_{\text{gap}}^{\text{GKS}} + \Delta_{\text{XC}}^{\text{den}}. \tag{12}$$

As $\Delta_{\text{XC}}^{\text{den}}$ depends uniquely on the density, it is the same within the two schemes and one obtains that

$$\varepsilon_{\text{gap}}^{\text{GKS}} - \varepsilon_{\text{gap}}^{\text{KS}} = \Delta_{\text{XC}}^{\text{orb}}. \tag{13}$$

The KS and GKS band gaps differ by the derivative discontinuity of the orbital-dependent XC energy functional $\Delta_{\text{XC}}^{\text{orb}}$. This confirms that the striking differences between the GKS and OEP-KS band gaps in Fig. 1 and the success of the GKS scheme is entirely due to the nonlocal treatment of the CHSX+ energy functional.

To illustrate these effects, we estimate $\Delta_{\text{XC}}^{\text{orb}}$ from first order perturbation theory in terms of the matrix elements of $v_{\text{XC}}^{\text{NL}}$ and $v_{\text{XC}}^{\text{orb}}$. By considering the GKS equation [Eq. (5)] as a perturbed KS equation with the perturbation $(v_{\text{XC}}^{\text{NL}} - v_{\text{XC}}^{\text{orb}})$ it follows that

$$\Delta_{\text{XC}}^{\text{orb}} \approx \langle \psi_{N+1}^{\text{KS}} | v_{\text{XC}}^{\text{NL}} - v_{\text{XC}}^{\text{orb}} | \psi_{N+1}^{\text{KS}} \rangle - \langle \psi_N^{\text{KS}} | v_{\text{XC}}^{\text{NL}} - v_{\text{XC}}^{\text{orb}} | \psi_N^{\text{KS}} \rangle. \tag{14}$$

From Fig. 1 we see that if we add to the OEP-KS gap the $\Delta_{\text{XC}}^{\text{orb}}$ contribution we almost recover the GKS value, with the residual difference due to higher order terms and to the different wave functions used in the two schemes.

Figure 1 and Eq. (13) represent the main message of this work: the poor performance of the OEP method for the band structure is not due to a poorly approximated correlation energy, but it is mainly due to the local treatment of the spatial dependence in the XC potential. In fact, as shown in Eq. (14), the spatial dependence in the XC potential induces an additional term in the GKS gap expression that well approximates the difference between the KS and the experimental gap.

Hence, the GKS scheme can be recommended as an efficient method for predicting the electronic structure in extended systems. Unlike the OEP-KS that requires an *a posteriori* calculation of the derivative discontinuity for all bands and **k** points [very cumbersome and limited by the range of validity of first order perturbation theory in Eq. (5)], the GKS partition of the XC functional into an orbital- and density-dependent part can be designed to provide a quite satisfactory description of the full band structure. Indeed, within time-dependent DFT, the time-dependent step has to correct the KS eigenvalue gap, however at present no ap-
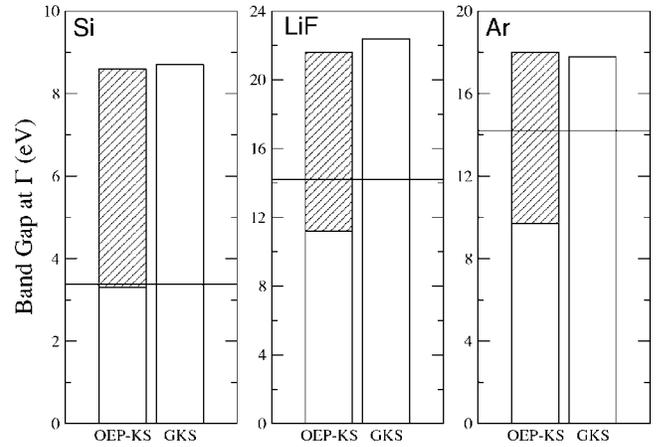

FIG. 2. Same as in Fig. 1 using an exchange-only energy functional (GKS reduces to the Hartree-Fock approximation). The striped bar corresponds to the first-order Δ_x contribution to the gap (see text). The continuous line indicates the experimental value (Refs. 30–32).

proximated XC functional can provide such correction. The limit of the GKS method remains the inclusion of dynamical effects that can be embodied considering the GKS as an approximation for the zero order Hamiltonian in a many-body self-energy calculation. Recently it has been shown that the GKS provides an excellent starting point—superior to standard KS—for *GW* calculations.[34]

Finally, the conclusions obtained for the CHSX+ approximation allow us to analyze in more detail the success of the exchange-only OEP-KS scheme for simple semiconductors (such as Si). In this limit the GKS reduces exactly to the Hartree-Fock approximation. In Fig. 2 the GKS, neglecting correlation, dramatically overestimates the experimental gap, while the OEP-KS band gap, following the same trend of Fig. 1, is much smaller than the GKS gap. The Δ_{XC} term reduces to the exchange-only Δ_x, given by Eq. (14) with $v_{\text{XC}}^{\text{NL}}$ replaced by v_x^{NL}, and correctly accounts for the difference between the OEP-KS and GKS gaps. For Si, the OEP-KS gap is in good agreement with the experiment. However for LiF and Ar, even if the OEP-KS band gaps are larger than in the case when correlation is included (Fig. 1), the band gaps remain 3 eV (LiF) and 4 eV (Ar) off the experimental gap. Consequently the good performance for Si of the OEP-KS scheme in the exchange-only limit is merely due to the balance of the neglected correlation that would close the gap, with the neglected Δ_x term that would open the gap to the huge Hartree-Fock value. The same approach that approximatively works for other semiconductors[7] fails in wide-gap insulators (such as LiF and Ar) where the delicate balance between correlation and exchange is more critical and cannot be neglected.

To summarize, we have found that the eigenvalue band gap changes dramatically when we use either a local OEP-KS or a nonlocal GKS potential, while keeping the same XC approximation. This trend is not specific for the approximations examined in this work. We observed it as well in the case of Hartree-Fock hybrid functionals and simple models for the screened exchange. As one would ex-

pect, the difference between the OEP-KS and the GKS band gap is due to the derivative discontinuity of the orbital-dependent XC energy that we have explicitly evaluated for Si, LiF, and Ar. This contribution makes the GKS band gap systematically larger than the OEP-KS gap and, in the case of the CHSX+ approximation, in very good agreement with the experimental results. These results give a clear and sound interpretation

of the OEP-KS scheme failure in terms of its intrinsic limit in describing the XC effects with a local potential.

---

*Present address: Université Catholique de Louvain, Unité de Physico-Chimie et de Physique des Matériaux, B-1348 Louvain-la-Neuve, Belgium.

[1] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[2] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982); J. P. Perdew and M. Levy, *ibid.* **51**, 1884 (1983).

[3] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983); Phys. Rev. B **32**, 3883 (1985).

[4] The proper framework to treat the band gap within KS is time-dependent DFT.

[5] On the other hand, good results can be obtained computing the band gap as differences of ground state energy. See, e.g., P. A. Schultz, Phys. Rev. Lett. **96**, 246401 (2006).

[6] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986); Phys. Rev. B **36**, 6497 (1987).

[7] M. Stadele, J. A. Majewski, P. Vogl, and A. Görling, Phys. Rev. Lett. **79**, 2089 (1997); M. Stadele, M. Moukara, J. A. Majewski, P. Vogl, and A. Görling, Phys. Rev. B **59**, 10031 (1999).

[8] R. T. Sharp and G. K. Horton, Phys. Rev. **90**, 317 (1953).

[9] J. D. Talman and W. F. Shadwick, Phys. Rev. A **14**, 36 (1976).

[10] M. E. Casida, Phys. Rev. A **51**, 2005 (1995).

[11] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).

[12] R. J. Magyar, A. Fleszar, and E. K. U. Gross, Phys. Rev. B **69**, 045111 (2004).

[13] M. Grüning, A. Marini, and A. Rubio, J. Chem. Phys. **124**, 154108 (2006).

[14] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy, Phys. Rev. B **53**, 3764 (1996).

[15] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[16] J. Heyd and G. E. Scuseria, J. Chem. Phys. **121**, 1187 (2004); J. Heyd, J. E. Peralta, G. E. Scuseria, and R. L. Martin, *ibid.* **123**, 174101 (2005); I. Prodan, G. E. Scuseria, and R. L. Martin, Phys. Rev. B **73**, 045104 (2006).

[17] D. M. Bylander and L. Kleinman, Phys. Rev. B **41**, 7868 (1990).

[18] We restrict the present discussion to spin unpolarized systems, the extension to spin polarization is straigthforward. Atomic units are used throughout the paper.

[19] L. Hedin, Phys. Rev. **139**, A796 (1965).

[20] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[21] To include the self-energy dynamical effects we renormalize *a posteriori* the eigenvalue band gaps by ∼0.9 in Fig. 1. We obtained the renormalization factor by comparing the CHSX and the *GW* (Ref. 13) OEP-KS eigenvalue band gaps.

[22] This approximation can be justified within MBPT with the Ψ functional formalism, see C. O. Almbladh, U. von Barth, and R. van Leeuwen, Int. J. Mod. Phys. B **13**, 535 (1999). Such a restriction can be released without affecting substantially our results as we have verified in Ref. 13.

[23] Computer code SELF, http://www.fisica.uniroma2.it/~self/

[24] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980); J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[25] Computer code ABINIT (http://www.abinit.org). See X. Gonze *et al.*, Comput. Mater. Sci. **25**, 478 (2002); Z. Kristallogr. - New Cryst. Struct. **220**, 558 (2005).

[26] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[27] M. Fuchs and M. Scheffler, Comput. Phys. Commun. **119**, 67 (1999).

[28] H. Monkhorst and J. Pack, Phys. Rev. B **13**, 5188 (1976).

[29] For the calculation of the nonlocal XC potential we used 259 reciprocal lattice vectors for Si, 331 for LiF (OEP), 821 for LiF (GKS), and 537 for Ar. In all cases we used 200 bands in the summation over conduction bands. In the sum over states of the OEP equation we used 100 bands for Si and 200 bands for LiF and Ar.

[30] *Semiconductors: Physics of Group IV Elements and III-V Compounds*, edited by O. Madelung, M. Schulz, and H. Weiss, Landolt-Börnstein. New Series, Group III, Part A Vol. 17 (Springer-Verlag, Berlin, 1982).

[31] M. Piacentini, D. W. Lynch, and C. G. Olson, Phys. Rev. B **13**, 5530 (1976).

[32] R. Haensel, G. Keitel, E. E. Koch, M. Skibowski, and P. Schreiber, Phys. Rev. Lett. **23**, 1160 (1969).

[33] In what follows we work within the ensemble approach that extends the domain of the definition of density functional to fractional number of particles and we consider only systems with an infinite number of electrons.

[34] F. Fuchs, J. Furthmüller, F. Bechstedt, M. Shishkin, and G. Kresse, cond-mat/0604447 (unpublished).