



Cite this: *Phys. Chem. Chem. Phys.*,
2017, 19, 15639

From the Kohn–Sham band gap to the fundamental gap in solids. An integer electron approach

E. J. Baerends 

It is often stated that the Kohn–Sham occupied–unoccupied gap in both molecules and solids is “wrong”. We argue that this is not a correct statement. The KS theory does not allow to interpret the exact KS HOMO–LUMO gap as the fundamental gap (difference $I - A$) of electron affinity (A) and ionization energy (I), twice the chemical hardness), from which it indeed differs, strongly in molecules and moderately in solids. The exact Kohn–Sham HOMO–LUMO gap in molecules is much below the fundamental gap and very close to the much smaller optical gap (first excitation energy), and LDA/GGA yield very similar gaps. In solids the situation is different: the excitation energy to delocalized excited states and the fundamental gap ($I - A$) are very similar, not so disparate as in molecules. Again the Kohn–Sham and LDA/GGA band gaps do not represent $I - A$ but are significantly smaller. However, the special properties of an extended system like a solid make it very easy to calculate the fundamental gap from the ground state (neutral system) band structure calculations entirely within a density functional framework. The correction Δ from the KS gap to the fundamental gap originates from the response part v^{resp} of the exchange–correlation potential and can be calculated very simply using an approximation to v^{resp} . This affords a calculation of the fundamental gap at the same level of accuracy as other properties of crystals at little extra cost beyond the ground state bandstructure calculation. The method is based on integer electron systems, fractional electron systems (an ensemble of N - and $(N + 1)$ -electron systems) and the derivative discontinuity are not invoked.

Received 2nd April 2017,
Accepted 30th May 2017

DOI: 10.1039/c7cp02123b

rsc.li/pccp

1. Introduction

The ionization energy I and electron affinity A are very important properties of electronic systems. The relation of the orbital energies of the Kohn–Sham (KS) one-electron model of density functional theory (DFT) to I and A is a source of considerable confusion. This is due to the fact that the relation between the orbital energies and the electron attachment and loss energies is different in different models. In the Hartree–Fock model the highest occupied molecular orbital (HOMO) energy is approximately equal to $-I$, and in the Kohn–Sham (KS) model it is exactly equal to $-I$. The LUMO energies are very different: approximately equal to $-A$ in HF but much lower in exact KS. The HOMO–LUMO gap is in HF approximately equal (in the frozen orbital Koopmans approximation) to the fundamental HF gap $I^{\text{HF}} - A^{\text{HF}}$, but in exact KS it is very close to the (much smaller) first excitation energy (the optical gap),^{1–8} see Fig. 1. How about standard DFT approximations (like local density approximation (LDA) and generalized gradient approximations (GGAs))?. They have both the HOMO and LUMO energies typically some 5 eV above the exact KS values,

see ref. 9 and references therein, and see Fig. 1. The situation is also different for molecules and solids, and these differences are again different for exact KS and for density functional approximations. For instance, in molecules the optical gap is much smaller than the fundamental gap $I - A$ (often only some 50%), and the exact KS HOMO–LUMO gap is close to the optical gap. In solids, on the other hand, the optical gap (for fully delocalized excitations) is often considered to be virtually identical to the fundamental gap, see Fig. 1. Making a distinction is then not necessary, which may confuse a molecular theoretician. The optical gap is obviously crucially important for techniques that rely on excitations of solids, such as in the harvesting of solar energy. We concentrate in this paper on the fundamental gap, and although the relevance for the optical gap should be evident, we do wish to caution that there are many aspects of excitations in solids that are not treated here. For instance, excitons have lower excitation energies than delocalized excitations, and may be responsible for an optical gap that is closer to the KS band gap. Moreover, the excited states that are responsible for the onset of “delocalized” photoabsorption are probably not delocalized over the entire crystal. The actual nature of these states will depend on experimental conditions such as the wavelength of the radiation, which may be much shorter than the dimensions of the crystal. The actual physics of the

Section Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands.
E-mail: e.j.baerends@vu.nl

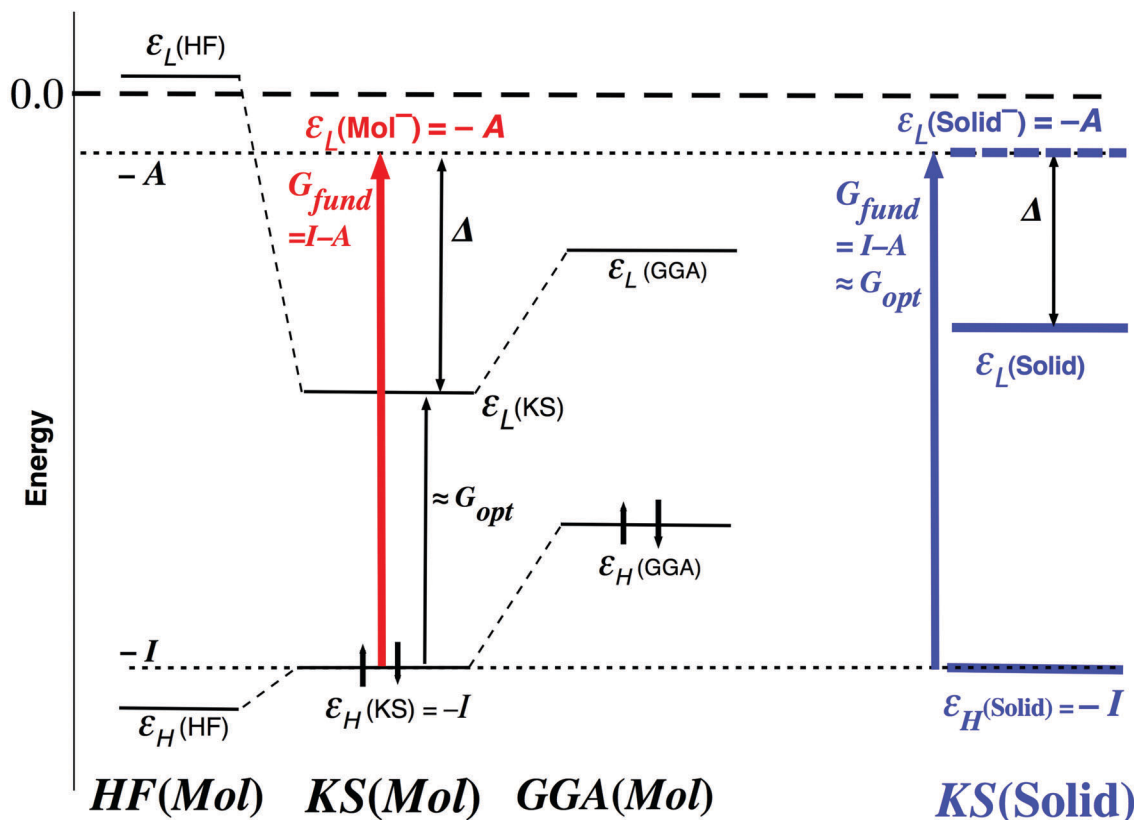


Fig. 1 Orbital energies of a molecule from the Hartree–Fock model (left panel), the exact Kohn–Sham model (second panel), and LDA or GGA approximations (third panel). The fundamental gap $I - A$ is indicated as the difference of the orbital energies of the highest occupied orbital (HOMO, H) in the neutral molecule and the singly occupied next higher orbital (L = H + 1) in the negative ion (the LUMO in the original neutral molecule). The corresponding orbital energies and fundamental gap for solids are pictured in the rightmost panel. The optical gap is in molecules approximated by the HOMO–LUMO gap in the neutral molecule, and is in solids (for fully delocalized excitations) practically equal to the fundamental gap. Since in many (small) molecules the HF LUMO is not bound, the HF LUMO has been positioned above the energy zero.

photoabsorption process in a macroscopic crystal is a topic far beyond the scope of the current paper.

Nevertheless, the fundamental gap, or just, somewhat indiscriminately, “the gap”, is considered the property of prime importance, and the “failure” of the KS band gap to reproduce the fundamental gap is considered a real disadvantage of density functional band structure calculations. Hybrid functionals confuse this picture somewhat, since by varying the amount of exact exchange from 0% to 100% (full Hartree–Fock) one moves the LUMO up from its “optical” value in molecules to the Hartree–Fock like value of $-A$, so hybrids change in molecules the gap from optical gap to fundamental gap. (This holds if one uses the nonlocal exchange operator, as generated by optimization of the hybrid total energy, which is often called generalized Kohn–Sham (GKS).) In the same way the band gap in solids can be “opened up” by allowing exact exchange in the total energy and one-electron Hamiltonian. The percentage of exact exchange can be determined empirically (not from first principles). This has led to rather successful methods of obtaining fundamental gaps from band structure calculations with hybrid functionals.^{10,11} We will not be using hybrid functionals, but we will deal with the question if it is possible to obtain in solids an accurate estimate of the fundamental gap

from density functional approximations that approximate the exact KS potential, which is local.

The orbital energy spectrum and the nature of the KS orbitals are determined by the potential in the one-electron Schrödinger equation they obey, the Kohn–Sham equation. In order to analyze this potential we use the partitioning of v_{xc} ,^{12,13} which naturally emerges from its definition in the Kohn–Sham (KS) theory as a functional derivative of E_{xc} with respect to the electron density $\rho(r)$

$$v_{xc}(r_1) = \frac{\delta E_{xc}[\rho]}{\delta \rho(r_1)} = \frac{\delta}{\delta \rho(r_1)} \left[\frac{1}{2} \int \frac{\rho(r_2)[\bar{g}(r_2, r_3) - 1]\rho(r_3)}{|r_2 - r_3|} dr_2 dr_3 \right]. \quad (1)$$

In (1) E_{xc} is expressed through the pair-correlation function $\bar{g}(r_2, r_3)$ integrated over the coupling parameter λ of the electron–electron interaction λ/r_{12} (coupling constant integrated quantities are denoted with an overbar). It determines the exchange–correlation hole $\bar{\rho}_{xc}^{hole}(r_2|r_1) = -[1 - \bar{g}(r_1, r_2)]\rho(r_2)$ that an electron at position r_1 builds in the total electron density at r_2 , so that it “feels” a Coulomb potential from the total charge density $\rho(r_2)$ which is diminished in its neighborhood by the

exchange–correlation hole with potential

$$\bar{v}_{xc}^{\text{hole}}(\mathbf{r}_1) = - \int \frac{[1 - \bar{g}(\mathbf{r}_1, \mathbf{r}_2)] \rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2, \quad (2)$$

$$E_{xc}[\rho] = \frac{1}{2} \int \rho(\mathbf{r}_1) \bar{v}_{xc}^{\text{hole}}(\mathbf{r}_1) d\mathbf{r}_1 = \int \rho(\mathbf{r}_1) \varepsilon_{xc}(\mathbf{r}_1) d\mathbf{r}_1 \quad (3)$$

Taking the derivative of eqn (1) it is clear that the exchange–correlation potential consists of two parts, the hole potential and the response potential,

$$v_{xc}(\mathbf{r}) = \bar{v}_{xc}^{\text{hole}}(\mathbf{r}) + \bar{v}_{xc}^{\text{resp}}(\mathbf{r}) \quad (4)$$

$$\bar{v}_{xc}^{\text{resp}}(\mathbf{r}_1) = \frac{1}{2} \int \frac{\rho(\mathbf{r}_2) \rho(\mathbf{r}_3) \delta \bar{g}(\mathbf{r}_2, \mathbf{r}_3)}{|\mathbf{r}_2 - \mathbf{r}_3|} \frac{\delta \rho(\mathbf{r}_1)}{\delta \rho(\mathbf{r}_1)} d\mathbf{r}_2 d\mathbf{r}_3, \quad (5)$$

If one does not use the coupling constant integration, but uses the kinetic correlation energy $T - T_s$ and the electron–electron exchange–correlation energy W_{xc} as separate terms in E_{xc} , the partitioning of v_{xc} is a bit different,

$$\begin{aligned} W_{xc} &= \frac{1}{2} \int \frac{\Gamma(\mathbf{r}_1, \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 - \frac{1}{2} \int \frac{\rho(\mathbf{r}_1) \rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \\ &= \frac{1}{2} \int \frac{\rho(\mathbf{r}_1) [g(\mathbf{r}_2, \mathbf{r}_2) - 1] \rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \\ &= \frac{1}{2} \int \rho(\mathbf{r}_1) v_{xc}^{\text{hole}}(\mathbf{r}_1) d\mathbf{r}_1 \end{aligned} \quad (6)$$

$$\begin{aligned} E_{xc} &= T - T_s + W_{xc} = T_c + W_{xc} \\ &= \int \rho(\mathbf{r}_1) v_{c,\text{kin}}(\mathbf{r}_1) d\mathbf{r}_1 + \frac{1}{2} \int \rho(\mathbf{r}_1) v_{xc}^{\text{hole}}(\mathbf{r}_1) d\mathbf{r}_1 \end{aligned} \quad (7)$$

$$v_{xc}(\mathbf{r}_1) = \frac{\delta E_{xc}}{\delta \rho(\mathbf{r}_1)} = v_{xc}^{\text{hole}}(\mathbf{r}_1) + v_{c,\text{kin}}(\mathbf{r}_1) + v_{xc}^{\text{resp}}(\mathbf{r}_1) \quad (8)$$

$$\begin{aligned} v_{xc}^{\text{resp}}(\mathbf{r}_1) &= \int \rho(\mathbf{r}_2) \frac{\delta v_{c,\text{kin}}(\mathbf{r}_2)}{\delta \rho(\mathbf{r}_1)} d\mathbf{r}_2 \\ &\quad + \frac{1}{2} \int \frac{\rho(\mathbf{r}_2) \rho(\mathbf{r}_3) \delta \bar{g}(\mathbf{r}_2, \mathbf{r}_3)}{|\mathbf{r}_2 - \mathbf{r}_3|} \frac{\delta \rho(\mathbf{r}_1)}{\delta \rho(\mathbf{r}_1)} d\mathbf{r}_2 d\mathbf{r}_3 \end{aligned} \quad (9)$$

The contribution of $v_{c,\text{kin}}$ to the potential and its response part is only important in special cases (such as at the bond midpoint of a dissociating molecule^{12,14–21}) and given the approximations we will be making (such as exchange-only approximation for the response potential) we can consider it negligible. We will henceforth use the straightforward expressions (6)–(9) (without coupling constant integration).

The hole potential may be called the “energy part” of the potential, since it is just (twice) the xc energy density, while the response potential^{12,22–26} reflects the response of the pair-correlation function to density variation. If one has an approximate functional that gives a good representation of the hole potential, one may expect decent results for the energy. This is how in fact the LDA and GGA derivations have started: with general considerations about a good modeling of the hole potential as function of the density. Even if such modeling gives reasonable values of the hole potential at positions \mathbf{r} ,

and therefore of the energy, it may still be that the analytic form of the density dependence is deficient, and the derivative, being sensitive to this analytical dependence, may be erroneous. It has recently been demonstrated that that actually is the case with the response part of the LDA and GGA potentials, which is why the orbital energies of these density functional approximations (DFAs) are in error by some 5 eV.⁹ The response potential plays generally an important role in the calculation of the orbital energies.^{9,25,27,28} Therefore it is crucially important in the present work. We recall that an exact representation of this potential can be given in terms of the ground state wavefunction Ψ_0^N and the KS ground state determinantal wavefunction $\Psi_{s,0}^N$,^{12,13} from which can be derived an expression in terms of the Dyson orbitals $\{d_i\}$ of the ionized states Ψ_i^{N-1} and the corresponding ionization energies plus the KS orbitals $\{\psi_j^s\}$ and their orbital energies^{25,29} (see Appendix A, eqn (A17) and (A18))

$$\begin{aligned} v_{xc}^{\text{resp}}(1) &= \sum_i (-I_0 + I_i) \frac{|d_i(1)|^2}{\rho(1)} - \sum_j (\varepsilon_N^s - \varepsilon_j^s) \frac{|\psi_j^s(1)|^2}{\rho(1)} \\ &= \sum_i \frac{|d_i(1)|^2}{\rho(1)} (I_i) - \sum_j \frac{|\psi_j^s(1)|^2}{\rho(1)} (-\varepsilon_j^s). \end{aligned} \quad (10)$$

where the second line follows since the Kohn–Sham HOMO orbital energy is exactly equal to the first ionization potential, $\varepsilon_H^s = -I_0$, and both the sum over (all) Dyson orbitals squared and over the occupied Kohn–Sham orbitals squared are equal to the total density, $\sum_i |d_i(1)|^2 = \sum_j |\psi_j^s(1)|^2 = \rho(1)$. (The sums run over spin orbitals, H denotes the highest occupied one in the neutral system, *i.e.* $H = N$.) The Dyson orbitals with large amplitude (Dyson orbitals are not normalized) resemble very closely the KS orbitals (or the HF orbitals).²⁵ This shows that the response potential will have, in atoms, a step like behavior where in each atomic shell, with atomic orbitals that build up the electron density of the shell, the step height will be determined by the difference in ionization energy and the KS orbital energy. This has been observed in calculations.^{22,28} In the region of the HOMO, where the HOMO KS orbital energy is equal to the first ionization potential, the step is zero. This potential has been approximated in the case of exact-exchange calculations (with a local potential, so optimized-effective-potential (OEP) or “ungeneralized Kohn–Sham”) by Krieger, Li and Iafrate,²⁷ and has been further approximated by Gritsenko *et al.*³⁰ as

$$v_{xc}^{\text{GLLB}}(\mathbf{r}_1) = \sum_{i=1}^H K \sqrt{\varepsilon_H - \varepsilon_i} \frac{|\psi_i(\mathbf{r}_1)|^2}{\rho(\mathbf{r}_1)}, \quad (11)$$

It can be shown that in the electron gas (EG) this is exactly the response part of the exchange potential with $K = 0.38$. Applying it in finite systems may be considered to be the EG approximation to the exchange response potential (it is not LDA since it does not work with just the local density). But the important aspect of this approximation is its reproduction of the step-like behavior the exact response part of the potential

exhibits in atoms and molecules.²⁸ We have argued that the traditional LDA exchange potential has a good hole part, which is even improved by Becke's gradient correction,⁹ but its response part (obtained by the functional derivative of the LDA exchange energy density) is very deficient in finite systems, leading to the poor orbital energies (upshift by *ca.* 5 eV). Replacing this deficient response part by the GLLB one of eqn (11) reduces the error in the orbital energies of atoms and molecules by about one order of magnitude.⁹ The hole part of the potential can still be obtained from the LDA or GGA energy density, see eqn (3). The exchange hole potential of the LDA model plus the Becke correction has been used^{9,30} in atomic and molecular calculations, while in ref. 9 also the LDA correlation hole potential (in VWN parametrization) has been employed. In solids the exchange–correlation energy density of the PBEsol functional has been used,³¹ $(1/2)v_{xc}^{hole}(r) \approx \epsilon_{xc}^{PBEsol}(r)$. In combination with the GLLB exchange response potential, this has been denoted as the GLLB-sc model.³¹ This is the model that will be used in applications in this paper.

In the present paper we focus on the application of a suitable hole potential and the GLLB response potential in solids. In the first place it will be demonstrated that this affords a simple and cheap calculation of the fundamental gap in solids. In Section III the physics of ionization and electron attachment in solids is discussed, showing that the KS gap (reproduced reasonably well with LDA/GGA calculations) is not an approximation of the fundamental gap. A detailed argument is given that the (exact) correction Δ to the (exact) KS gap, which is needed to obtain the fundamental gap, is given by the expectation value of the exact response part eqn (10) of the potential. In Section IV it is shown that the simple GLLB approximation of eqn (11) makes it possible to calculate the discrepancy Δ between fundamental and KS gap in a simple manner, just as a correction after the regular band structure calculation. The possibility for this cheap and efficient calculation of Δ has been recognized by Kuisma and Rantala and coworkers³¹ and it has been investigated and applied by Thygesen and Jacobsen and coworkers.^{32–37} It leads to excellent approximations of the fundamental gap.

Our justification and explanation (Section III) work entirely with integer electron systems, as in fact is done in the work of Sham and Schlüter^{38,39} who also evaluate the difference in the KS potentials of $(N + 1)$ - and N -electron systems, $v_{xc}(N + 1) - v_{xc}(N)$, *cf.* Godby, Schlüter, Sham.^{40–42} These authors use many-body techniques (the GW method) to evaluate this difference. On the other hand, an explanation of the difference $v_{xc}(N + 1) - v_{xc}(N)$ is often sought from the behavior of the total energy at fractional electron numbers and the derivative discontinuity of the total energy at integer electron number^{43,44} (“fractional electrons” is a shorthand for the use of ensembles of states with different electron numbers; this ensemble approach⁴³ is receiving considerable interest in the literature^{45,46}). The present results show that the discrepancy between the KS band gap and the fundamental gap can be understood, and indeed be calculated, without any reference to fractional electron numbers or ensembles. This invites a closer examination of the concepts of

fractional electron numbers, the derivative of the energy with respect to electron number (and its discontinuity) as well as the derivative of the energy with respect to occupation numbers (Janak's theorem). This will be taken up in a follow-up paper.

II. Molecules: Kohn–Sham gap, optical gap and fundamental gap

We summarize in Fig. 1 the situation for the orbital energy spectra of the various models for a molecule. The Hartree–Fock model has the HOMO energy at the “frozen orbital” ionization energy. When the ion is described by the frozen ground state orbitals, the relaxation of the orbitals in the ion from their neutral ground state shapes to the orbitals in the ion is neglected (thus frozen orbital IP estimate too high). Also the difference in correlation energies in the ion and molecule is neglected in HF. As detailed in ref. 6 one may define the relaxation errors and the correlation errors in the cases of ionization and attachment as positive quantities, $\Delta_{\text{relax}}(I)$ and $\Delta_{\text{corr}}(I)$ for these energies in the positive ion, and $\Delta_{\text{relax}}(A)$ and $\Delta_{\text{corr}}(A)$ for these energies in the negative ion, and obtain

$$\begin{aligned}\epsilon_{\text{H}}^{\text{HF}} &= -I - \Delta_{\text{relax}}(I) + \Delta_{\text{corr}}(I) \\ \epsilon_{\text{L}}^{\text{HF}} &= -A + \Delta_{\text{relax}}(A) + \Delta_{\text{corr}}(A).\end{aligned}\quad (12)$$

This expresses the well known fact that the HF HOMO orbital energy is a reasonable approximation to the experimental ionization energy because the relaxation error and the correlation error partly cancel. In fact, the relaxation is usually about 1 eV larger, and the HF orbital energy in molecules is then about 1 eV more negative than $-I$. For the electron affinity the two errors work in the same direction, and therefore the HF LUMO orbital is typically much too high lying, often even at positive orbital energy, see ref. 6 for a discussion of this LUMO problem in HF. In the KS model the HOMO energy is at exactly minus the first ionization energy, $-I$, on account of well known asymptotic decay arguments. Unfortunately, the LDA and GGA approximations make a large systematic error in the occupied orbital energies, raising them by some 4–6 eV. This is not due to a self-interaction error, but to a too strongly repulsive response potential part of the LDA exchange potential.⁹ The low-lying virtual orbitals of valence character are raised by about the same amount, so the HOMO–LUMO gap of LDA/GGA is still close to the exact KS gap.^{6,41,47} There is considerable evidence that the exact KS gap is very close to the first excitation energy.^{1,3,6–8,48} The virtual orbitals are solutions in the same potential as the occupied orbitals, therefore physically they represent excited electrons in the non-interacting Kohn–Sham system. The crucial feature of the KS potential is the presence of the attractive potential of the exchange–correlation hole with -1 electron charge. Effectively electrons should move in the field of $N - 1$ other electrons, which means that the Hartree potential of the total electron density of N electrons has to be corrected by such a hole potential. For occupied orbitals this is usually referred to as the correction of the self-interaction error

in the Coulomb potential. For excited electrons the fact that they “see” only $(N - 1)$ other electrons is usually pictured in terms of the hole that is “left behind” when the electron-hole pair of the excitation is formed. The KS potential incorporates the stabilizing potential of the exchange–correlation hole, which accounts for this important hole potential effect on the excited electron, even if only approximately.

How important this stabilization of the LUMO and other virtual orbitals is, becomes clear when comparing to the LUMO energy of the Hartree–Fock model. In HF the virtual orbitals feel the repulsive potential of N electrons, there is no self-interaction correction. Physically the HF virtual orbitals describe added electrons, not excited electrons. The HF LUMO orbital energy is actually equal to the negative of the frozen orbital approximation to the HF electron affinity, $\epsilon_L^{\text{HF}} = -A^{\text{HF}}(\text{frozen orbital})$ (Koopmans theorem for electron addition), cf. eqn (12). For the vast majority of molecules the electron affinity is quite small, and the frozen orbital approximation of the affinity makes relaxation and correlation errors of the same sign, see eqn (12). These errors both lead to $\epsilon_L^{\text{HF}} = -A^{\text{HF}}(\text{frozen orbital})$ to be larger than $-A$, so ϵ_L^{HF} is very close to zero, and in many molecules, in particular small ones, there is not even a LUMO orbital with negative orbital energy. This explains the huge difference between the KS and the HF HOMO–LUMO gaps depicted in Fig. 1.

It is to be noted that the LUMO ϕ_L becomes the highest occupied orbital of the negative ion M^- , and its exact KS orbital energy must be at minus the ionization potential of M^- , which is $-A$ of the neutral M . [M stands for neutral molecule, M^- for its negative ion; C will stand for the neutral crystal, C^- for the crystal with one electron added. H is the N -th spinorbital (the HOMO in the neutral systems); L is the $(N + 1)$ -th spinorbital, the LUMO in the neutral systems and the HOMO in the negative ions. H and L are always equal to N and $(N + 1)$, independent of the electron count of the systems.] In the HF model, this level now benefits, as occupied orbital, from the stabilization of the self-interaction correction term in the exchange operator, and it becomes a realistic approximation to the ionization energy of the negative ion, i.e. of $-A$, because the frozen orbital approximation is not so poor any more, with counteracting relaxation and correlation errors, see eqn (12).

With any density functional approximation, the ionization energy and electron affinity can in principle be calculated straightforwardly in molecules by calculating the energies of the positive ion (M^+) and negative ion (M^-) in separate SCF calculations.⁴⁹ However, the erroneous upshift of the orbitals by ca. 5 eV in standard approximations often causes the occupied orbital in M^- (the former LUMO) to shift above zero, where its orbital energy and spatial extent become undefined. This problem for negative ions has nothing to do with DFT as such, but is simply a consequence of the error in the potentials that follow from standard DFT approximations. With correct potentials, negative ions are stable in DFT.⁵⁰

Within the pure KS context a Koopmans’ approximation to the shift Δ of ϵ_L to $-A$ is as easily obtained^{6,51} as it is from a HF or generalized KS calculation. Using $v_s(N)$ and $v_s(N + 1)$ for the

KS potentials of the N and $N + 1$ systems resp., and writing $\phi_L(M^-) = \phi_L^-$ and $\phi_L(M) = \phi_L$ it is clear that

$$\Delta = \epsilon_L^- - \epsilon_L = \langle \phi_L^- | \hat{T} + v_s(N + 1) | \phi_L^- \rangle - \langle \phi_L | \hat{T} + v_s(N) | \phi_L \rangle \quad (13)$$

This is an exact expression for Δ . Ignoring electron relaxation effects in the anion (Koopmans’ frozen orbital approximation), i.e. with $\phi_L(M^-) = \phi_L(M)$ and $\rho(M^-) = \rho(M) + |\phi_L|^2$, we can approximate this quantity as

$$\begin{aligned} \Delta &\approx \langle \phi_L | v_{\text{Coul}}(|\phi_L|^2) + v_{\text{xc}}(\rho^M + |\phi_L|^2) - v_{\text{xc}}(\rho^M) | \phi_L \rangle \\ &= \langle \phi_L \phi_L | \phi_L \phi_L \rangle + \langle \phi_L | v_{\text{xc}}(\rho^M + |\phi_L|^2) - v_{\text{xc}}(\rho^M) | \phi_L \rangle \end{aligned} \quad (14)$$

The frozen orbital approximation may be more severe here than in the case of Hartree–Fock, since the unoccupied ϕ_L is not so diffuse as would be appropriate for the negative ion. In large molecules, which are typically of interest in CT excitations, this may not be so serious. We note that the potentials $v_{\text{xc}}(\rho(M^-))$ and $v_{\text{xc}}(\rho(M))$ both incorporate the attraction of an electron hole with -1 charge. The largest part of the upshift Δ comes from the self-repulsion Coulomb integral $\langle \phi_L \phi_L | \phi_L \phi_L \rangle$. Chai and Chen⁵² have shown that (14) is actually the first term in a perturbation expansion which may be carried further for higher accuracy, although the first term already yields very decent approximation.

III. Solids: Kohn–Sham gap, optical gap and fundamental gap

The KS gap in solids is known to be smaller than the fundamental gap $I - A$. The physical argument why the exact KS orbital energy of the lowest unoccupied orbital (KS LUMO or bottom of the conduction band) in the neutral system will not be equal to (close to) $-A$, also in an extended solid, is the same as in molecules. As has been pointed out in ref. 6, the $v_{\text{xc}}^{\text{hole}}$ part of the KS potential exerts a significant stabilizing effect on the LUMO level, since the exchange–correlation hole in solids is localized. It is not infinitely extended (as the orbital self-interaction part of this hole, $-|\phi_L|^2$ would be). On the other hand, in a fully delocalized excitation there is not a localized hole that is left behind, and there is not a discernable stabilizing effect on the one-electron state of the excited electron. The one-electron energies of the KS model therefore need not match delocalized excitation energies very well. For the same reason one should expect the KS LUMO (bottom of the conduction band) to be lower lying than $-A$,⁶ see Section IIIA. Density functional approximations (DFA) like LDA and GGA, which also employ a local exchange–correlation model hole, approximate the exact KS orbital energy gap in solids very well,^{6,31,40,41,47} as they do in molecules.⁷ So the orbital energy gap of LDA/GGA being smaller than the optical/fundamental gap is not considered to be a deficiency of the DFAs, but to be an inherent property of the KS model, see Fig. 2.

It is the purpose of the present work to demonstrate that the special properties of crystals, notably their macroscopic dimensions, make it rather easy to calculate the fundamental

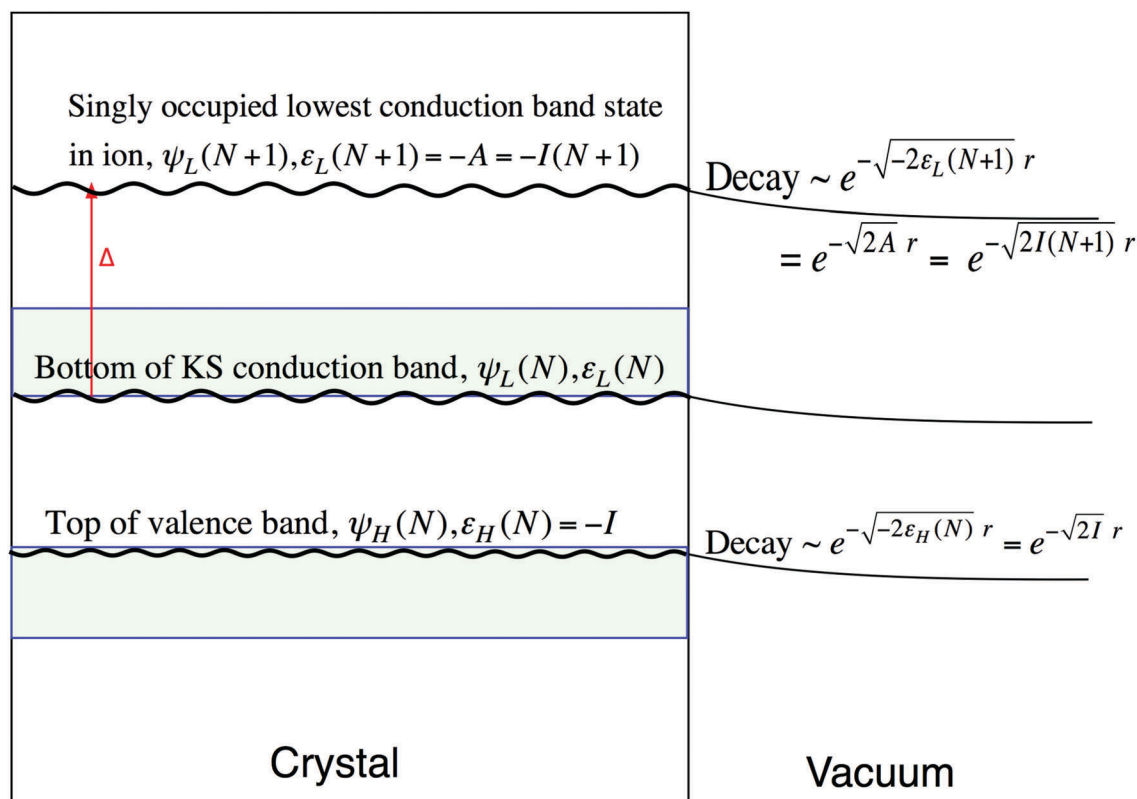


Fig. 2 Schematic picture of the effect of the behavior of the KS one-electron states at the top of the valence band and at the bottom of the conduction band. The latter is also given in the $(N + 1)$ -electron system, where one electron has been added to the crystal and the one-electron states have moved up by Δ . The behavior inside the crystal is sketched as well as the decay outside the crystal.

gap (the “correction” Δ to the KS orbital energy gap) from just the usual ground state DFA band structure calculations. In Section IIIA we will give the physical argument why the KS band gap is not the fundamental gap, using the partitioning of the exchange–correlation potential v_{xc} of eqn (4) and (8). In the next Section (IIIB) we will argue that the correction Δ from KS band gap to fundamental gap must be given by the difference of the response potential (defined *e.g.* in eqn (5)) of the crystal with one electron added (C^-) minus the neutral crystal (C) response potential. In the last Section (IIIC) we will derive from the known explicit exact expression for the response potential, eqn (10), that indeed this potential exhibits a uniform upshift (over the crystal) of precisely Δ when going from C to C^- .

A. Why the Kohn–Sham LUMO level is below $-A$

In order to understand the band gap problem we need to understand how a one-electron model like Kohn–Sham can be related to the actual processes of ionization and electron attachment in a crystal. When treating ionization and electron attachment we have to realize that these processes can only be understood when we consider a crystal as a real finite piece of crystalline material, which may have macroscopic dimensions (1 cm³ say). Therefore the number of unit cells will be large (in the order of Avogadro’s number), but still finite. The (very large) outside world is an integral part of the system we are studying, because it is there that the escaping electron (with ionization) is

observed and *e.g.* its kinetic energy measured (in a photoemission experiment). The presence and role of the outside world are as crucial to understand the properties of crystals as they are for molecules, see below. To what extent can the KS one-electron model be used to get a first approximation to the processes (excitation, ionization, electron attachment) we want to study? The one-electron states (orbitals) can be approximated in the interior of the crystal, where boundary effects are negligible, by using the standard assumption of band structure calculations that the crystal is infinite (or apply cyclic boundary conditions). In this way the electron density inside the crystal can be approximated at the same level of accuracy as in molecules. In order to determine the orbital energies, we fix the gauge of the KS potential by requiring that it goes to zero at infinity. (In an infinite crystal we cannot fix the gauge of the potential and only the orbital energy differences are meaningful, so here we need the vacuum outside the crystal, *cf. e.g.* ref. 53.) Just as in molecules, the decay of the electron density outside the crystal will be determined by the slowest decaying orbital density, which is for the highest-energy occupied orbital (the top of the valence band). In the same way as for molecules (*i.e.* for all finite systems), the asymptotic decay of the density into the vacuum is governed by the ionization energy I , $\rho(r) \sim e^{-2\sqrt{2I}r}$. This dictates the highest occupied orbital energy to be $\epsilon_H = -I$ in order to have the correct asymptotic decay of the highest occupied orbital density $|\psi_H(r)|^2 \sim e^{-2\sqrt{-2\epsilon_H}r} = e^{-2\sqrt{2I}r}$.

If we add an electron to the system, the density decay will be in keeping with the ionization energy of the negative ion, *i.e.* the electron affinity A of the neutral system, $|\psi_L(N+1)(\mathbf{r})|^2 \sim e^{-2\sqrt{2A}r}$. [We recall that the highest occupied spinorbital of the neutral system is labeled H , and the lowest unoccupied spinorbital with $L = H + 1$. H and L continue to denote the same levels in the negative ion, where orbital L becomes occupied.] Given the extended nature of solids, it has initially been thought that addition of one electron could only make an infinitesimal difference to the KS potential, so the orbital energies should stay the same and one should also have for the N -electron system $\varepsilon_L(N) = -A$.^{54,55} This was in contradiction with results from approximations (LDA/GGA). It was then realized by Perdew and Levy⁴⁴ and Sham and Schlüter³⁸ that even addition of one electron could cause a finite shift Δ of the orbital energies, yielding the relation $\varepsilon_L(N+1) = \varepsilon_L(N) + \Delta = -A$, as depicted in Fig. 2.

There is a simple physical reason why the KS LUMO level of the neutral (N -electron) system cannot be expected to be at $-A$. The KS potential contains the exchange–correlation potential v_{xc} , which has as most important part the potential of the exchange–correlation hole charge density, v_{xc}^{hole} . In the present context it is important that the hole charge density, which integrates to -1 electron, is rather localized around the position \mathbf{r}_1 where the potential is required (the reference position), at least when that position is in the bulk region of the molecule or solid,^{56–58} and not in the asymptotic region.⁹ In the Hartree–Fock model the orbital self-interaction part $-|\psi_k(\mathbf{r}_2)|^2$ (independent of the reference position \mathbf{r}_1 where the hole potential is required) of the hole for a specific orbital ψ_k , even though it determines the total charge of -1 of the hole, is not determining its shape. If there are many occupied orbitals, the remaining terms that make up the exchange hole charge will make the hole much more localized around the reference position \mathbf{r}_1 . It has been emphasized by Luken⁵⁹ that this can be derived from an orbital localization argument, where one can see that when a unitary rotation is

effected on the occupied orbital set, the localized orbital with largest amplitude at the reference position \mathbf{r}_1 determines the shape of the hole. In for instance “delocalized” π -electron systems (*e.g.* benzene) the exchange hole does not at all look like a delocalized -1 charge over the π system, which would be in keeping with the orbital self-interaction part of the hole, but is almost completely confined to one atomic region.⁶⁰ The correlation hole has zero charge, but its shape (negative around \mathbf{r}_1 , positive further away) further enhances the localized nature of the total xc hole.⁶⁰ This is in perfect agreement with what is known for solids: in a solid like Si the exchange–correlation hole is approximately confined to one unit cell.⁵⁸ The rather localized xc hole makes it very different from the infinitely extended “hole” $-|\phi_i(\mathbf{r})|^2$ that arises as a consequence of the creation of an electron–hole pair in the delocalized excitation $\phi_i \rightarrow \phi_a$. This is why we may physically not expect the orbital energy difference to match approximately the energy of a delocalized excitation. In contrast, in small molecules where the hole of the electron–hole pair of the excitation has similar extent as the exchange–correlation hole, the similar stabilizing potentials cause the orbital energy difference to be comparable to (actually quite close to) the excitation energy.^{6,7}

B. v^{resp} as origin of the constant upshift of the KS potential from neutral crystal C to negative ion C^-

Let us consider a crystal to which an electron is added. As depicted in Fig. 3 we assume that in the interior of the crystal the behavior of the one-electron states and the density can be obtained from calculations with periodic boundary conditions (band structure calculations). Band structure calculations are based on the assumption that for unit cells deep inside the crystal the boundary effects and the outside world have negligible influence. The effect of the addition of one electron to the crystal on the electron density is totally negligible per unit cell, so in the unit cells in the interior of the crystal the electron

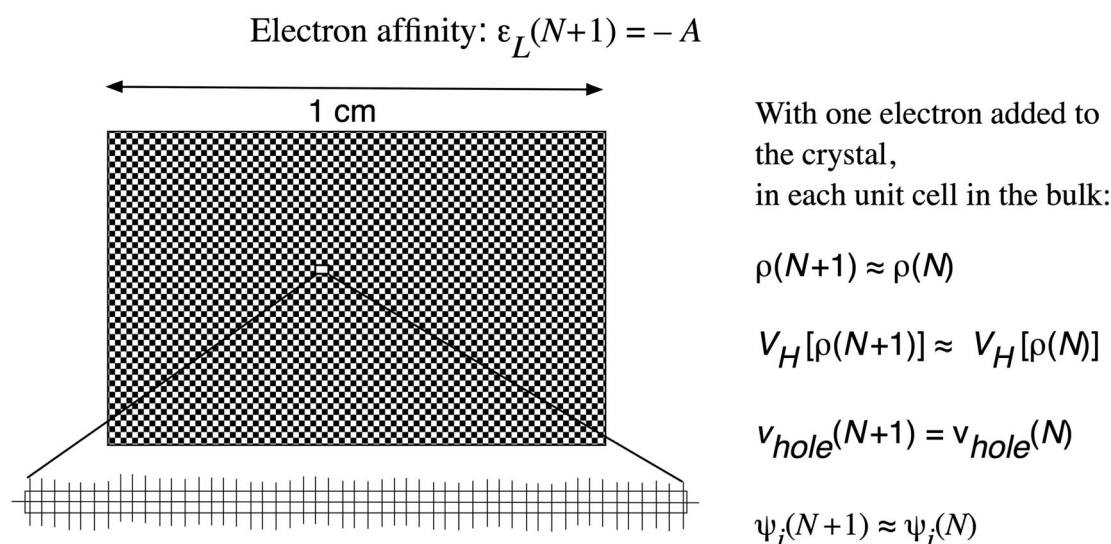


Fig. 3 Schematic picture of the effect of addition of one electron to a crystal, creating the $(N+1)$ -electron system, on the electron density, Hartree potential and one-electron states in unit cells inside the crystal.

density will stay practically the same. This immediately suggests that the KS potential of the negative system will be unchanged or be changed by a uniform constant over the crystal. A local change of shape of the potential (in each unit cell) is, with the requirement of unchanging density, forbidden by the Hohenberg–Kohn (HK) theorem. The HK theorem states that $\int \Delta \rho \Delta v < 0$ for any change Δv other than a constant,

$$\int \Delta v \Delta \rho \, dr < 0 \quad (15)$$

telling us that if there would be a local change Δv of the potential, over a limited region, the density would have to exhibit a change locally (over the same region).^{50,61} Given that the added electron makes an essentially zero contribution to the density in a unit cell, any change in the density must result from a change in the density of the underlying N -electron system, $\Delta \rho(N)$. If the KS potential would exhibit a true (non-constant) change over some region of the unit cell, the HK theorem stipulates that over that region we must have a $\Delta \rho(N) \neq 0$ in order to obey the HK inequality $\int \Delta \rho \Delta v < 0$. The fact that $\Delta \rho(N) = 0$ leaves only the possibility that the KS potential shifts by a constant. The KS potential being the same in all unit cells, the constant must be uniform over the crystal (except of course in the border region and outside, where it will always, by definition of our gauge, go asymptotically to zero).

So the KS potential of C^- may shift by a constant, and indeed it must. The situation is pictured in Fig. 2. The density decay outside a neutral system of any (finite) size (molecule or crystal) is governed by the ionization energy of the system. So the highest occupied orbital (top of the valence band) decays like $\exp[-\sqrt{-2\varepsilon_H(N)}r] = \exp[-\sqrt{2I}r]$. The decay of the lowest unoccupied state (at the bottom of the conduction band) will be $\exp[-\sqrt{-2\varepsilon_L(N)}r]$. In the negative ion, the density decay will be governed by the ionization energy of that system, A , which must be the density decay of the highest occupied orbital in that system, the state at the bottom of the conduction band: $\exp[-\sqrt{-2\varepsilon_L(N+1)}r] = \exp[-\sqrt{2A}r]$ (see Fig. 2).

So there must be a uniform upshift of the KS potential over the bulk of the crystal with magnitude

$$\begin{aligned} \Delta &= \varepsilon_L(N+1) - \varepsilon_L(N) = I - A - g^{\text{KS}} \\ g^{\text{KS}} &\equiv \varepsilon_L(N) - \varepsilon_H(N) \end{aligned} \quad (16)$$

where g_{KS} is the KS gap. The upshift of the potential does not extend into the asymptotic region, the asymptotic value of the potential is kept at zero at infinity in order to have a meaningful comparison of the potentials in the various situations. Because the expectation value $\langle \phi_L(N+1) | - (1/2)\nabla^2 + v_s(N+1) | \phi_L(N+1) \rangle$ is made up over the bulk of the crystal, the asymptotic region where the potential is not upshifted gives negligible contribution. So all orbital energies will be upshifted by precisely the same amount, $\varepsilon_i(N+1) - \varepsilon_i(N) = \Delta \forall i$. It is to be noted that the infinitely extended vacuum outside the crystal is crucial to define the magnitude of the constant upshift of the potential inside the crystal.

We have not invoked here the idea of a derivative discontinuity of the energy and the discontinuous jump of the KS potential⁴⁴

by a constant for an infinitesimal (or any fractional) electron addition, but we are considering the constant upshift of the Kohn–Sham potential when one full electron is added to a crystal. The difference Δ between the KS potentials of the two integer electron systems was called a “discontinuity” in the KS potential (not a derivative discontinuity) by Sham and Schlüter,^{38,39} since also the addition of a full electron to a solid implies only a minute change in electron density per unit cell. It has been numerically established by Godby, Schlüter and Sham^{40–42} using the GW method. We will discuss a much cheaper purely density functional method to obtain the correction Δ . In order to describe the upshift of the orbital energies we consider the nature of the energy spectrum of the one-electron KS states (orbitals) in a solid. They are all solutions in the same local potential

$$\begin{aligned} v_s &= v_{\text{nuc}} + v_H + v_{\text{xc}} \\ v_{\text{xc}} &= v_{\text{xc}}^{\text{hole}} + v_{\text{c,kin}} + v^{\text{resp}} \end{aligned} \quad (8)$$

The nuclear potential (the external potential in our systems) and the Hartree potential of the total electronic density are trivial (they are common to virtually all one-electron models, in particular also Hartree–Fock). Upon electron addition, the nuclear potential is evidently unchanged, and with $\Delta \rho(N) = 0$, also the Hartree potential will be unchanged inside the crystal, $v_H[\rho(N+1)] = v_H[\rho(N)]$. The exchange–correlation hole potential $v_{\text{xc}}^{\text{hole}}(\mathbf{r}_1)$ and the kinetic correlation potential $v_{\text{c,kin}}(\mathbf{r}_1)$ depend on the correlation as embodied in the two-particle (and one-particle) density matrices of the interacting systems. For any reference position \mathbf{r}_1 inside the crystal they will be unchanged, since the conditional probability to find another electron in the vicinity of the reference electron will not be changed by the one electron that is added to the enormous number already present. All these “physical” contributions to the KS potential do not change. The only part of the KS potential that can change inside the crystal is then v^{resp} ,^{12,24–26,29} and if $-A - \varepsilon_L = \Delta \neq 0$, it must be the response part v^{resp} of the KS potential that changes by the constant Δ over the whole crystal. With a constant change Δ , the KS one-electron states will not change shape, $\psi_i(N+1) = \psi_i(N)$ consistent with the unchanging density.

C. Δ from exact v^{resp} of eqn (10)

We now show that the constant upshift of the KS potential in the negative ion is exactly provided by the response potential. For this purpose the breakdown of the response potential as a difference of two terms, related to the exact wavefunction and the KS determinantal wavefunction, respectively, can be used (see Appendix A),

$$v^{\text{resp}}(1) = v^{N-1}(1) - v_s^{N-1}(1) \quad (17)$$

$$v^{N-1}(1) = \sum_i (-I_0 + I_i) \frac{|d_i(1)|^2}{\rho(1)} \quad (18)$$

$$v_s^{N-1}(1) = \sum_i (\varepsilon_H - \varepsilon_i) \frac{|\psi_i^s(1)|^2}{\rho(1)} \quad (19)$$

(we sum over spinorbitals, so ε_N is the highest occupied orbital energy, *etc.*). Here I_i are the ionization energies of the interacting system (I_0 the first one), and d_i the corresponding Dyson orbitals (see eqn (A18)), while ψ_i^s are the KS orbitals and ε_i the corresponding orbital energies. We are now interested in the change in the response potential when going from the neutral crystal (denoted C) to the negative ion C^- , which can be partitioned into exact wavefunction and KS wavefunction parts,

$$\begin{aligned}\Delta v^{\text{resp}} &= (v^{N-1}[C^-] - v_s^{N-1}[C^-]) - (v^{N-1}[C] - v_s^{N-1}[C]) \\ &= (v^{N-1}[C^-] - v^{N-1}[C]) - (v_s^{N-1}[C^-] - v_s^{N-1}[C]) \\ &= \Delta v^{N-1} - \Delta v_s^{N-1}\end{aligned}\quad (20)$$

The potentials $v^{N-1}[C^-]$ and $v_s^{N-1}[C^-]$ of the negative ion can be written according to eqn (17) in the same way as for the neutral system, taking into account that the first ionization energy of C^- is equal to A and the KS $\varepsilon_i[C^-] = -A$. As for the KS part of Δv^{resp} , we are considering the case that all KS occupied orbital energies of C^- are upshifted by a constant with respect to those of C , $\varepsilon_i(C^-) = \varepsilon_i(C) + A$, so from eqn (19)

$$\begin{aligned}\Delta v_s^{N-1} &= v_s^{N-1}[C^-](1) - v_s^{N-1}[C](1) \\ &= \sum_i^L (-A - \varepsilon_i(C) - A) \frac{|\psi_i^s(C^-)(1)|^2}{\rho[C^-](1)} \\ &\quad - \sum_i^H (-I - \varepsilon_i(C)) \frac{|\psi_i^s(C)(1)|^2}{\rho[C](1)}\end{aligned}\quad (21)$$

The highest term in each summation (H for C and (L) for C^-) drops out, so that in both cases asymptotically the numerators $|\psi_{i<H}^s(C)(1)|^2$ and $|\psi_{i<L}^s(C^-)(1)|^2$ become negligible compared to the denominators $\rho(C)(1)$ and $\rho(C^-)(1)$ respectively. Asymptotically Δv_s^{N-1} of (21) will therefore go to zero. Inside the crystal we use that $|\psi_L^s(1)|^2$ is negligible at each point compared to $\rho[C](1)$, so $\rho(C^-) = \rho(C)$ and as a consequence $\psi_i^s(C^-) = \psi_i^s(C)$,

$$|\psi_L^s(1)|^2 \ll \rho[C](1) \rightarrow \rho[C^-](1) = \rho[C](1)$$

and

$$\frac{\sum_i^N |\psi_i^s(1)|^2}{\rho(1)} \approx 1 \quad \text{for } r_1 \in \text{crystal} \quad (22)$$

Substituting in (21) one obtains

$$\begin{aligned}\Delta v_s^{N-1}(1) &= (I - A - A) \sum_i^H \frac{|\psi_i^s(1)|^2}{\rho[C](1)} \\ &= (I - A - A) = g^{\text{KS}} \text{ for } r_1 \in \text{crystal}\end{aligned}\quad (23)$$

Turning now to the Δv^{N-1} part of the change in the response potential, we have from eqn (18) and (A18)

$$\Delta v^{N-1} = \sum_i (-A + \bar{I}_i) \frac{|\bar{d}_i(1)|^2}{\bar{\rho}(1)} - \sum_i (-I + I_i) \frac{|d_i(1)|^2}{\rho(1)} \quad (24)$$

where we denote the quantities for the negative ion with an overbar (\bar{I}_i , \bar{d}_i , $\bar{\rho}$). $I = I_0$ is the first ionization energy of the

neutral system. Of course the ionization energy A of the additional electron appears as the lowest ionization energy of C^- , $\bar{I}_0 = A$. The corresponding term in the summation for C^- drops out since $\bar{I}_0 = A$. We can now use that the other ionization energies \bar{I}_i of the ion C^- will be equal to the I_i of the neutral C . This would not be true at all for a small system like a molecule (of size of the order of 10 Bohr, say). It is a consequence of the macroscopic nature of the crystal C . This makes the interaction energy between the electron in the extended Bloch state at the bottom of the conduction band (LUMO) in C^- with the electron (or hole) in $|\psi_i|^2$ negligible.

We can picture the ionization of C^- arising from removing an electron from orbital ψ_i in the C^- system, yielding the neutral $C(\psi_i^{-1}\psi_L)$, in various ways. Suppose we take the route of first forming the neutral system C by ionizing the added electron of C^- from orbital ψ_L (energy rise $+A$). Next, ionization from ψ_i to $C^+(\psi_i^{-1})$ has a cost I_i . Finally, electron capture by the $C^+(\psi_i^{-1})$ ion of an electron in the LUMO to form $C(\psi_i^{-1}\psi_L)$ corresponds to an energy change (lowering) $-A_i$ (in principle the electron affinity A_i of $C^+(\psi_i^{-1})$ could be different from the affinity when there is not a hole in ψ_i , which is A)

$$\begin{aligned}\bar{I}_i &= E(C(\psi_i^{-1}\psi_L)) - E(C^-(\text{gr.st.})) \\ &= (E(C) + I_i - A_i) - (E(C) - A) = I_i + A - A_i = I_i\end{aligned}\quad (25)$$

For the last equality we have used that actually A_i is equal to the affinity of the neutral system C if C has macroscopic dimensions. Suppose C has radius 1 mm $\approx 10^7$ Bohr, and the positive hole charge of $C^+(\psi_i^{-1})$ is smeared out over the complete crystal, then the electrostatic energy gain of bringing an electron from infinity to the surface of the crystal is in the order of 10^{-7} a.u. This difference with the neutral crystal is totally negligible, so $A_i = A$. This results in $\bar{I}_i = I_i$. That would be very different if the system would have dimension 10 Bohr, as in the case of molecules. Of course there is in principle a difference between ionizing C^- and C . In the first case also the repulsion between the electron density $|\psi_i|^2$ and the electron density $|\psi_L|^2$ is released. But with macroscopic dimensions of these charge distributions, the effect is nil. Finally, the Dyson orbitals \bar{d}_i of C^- are expected to be very close to the Dyson orbitals of the neutral crystal C . This is in keeping with the argument that the ionisations of the C^- and C systems are the same (except for the first ionization of C^- , $\bar{I}_0 = A$). This can be further substantiated if we consider how close the Dyson orbitals are to the orbitals of a one-determinant approximation of the wavefunction, for which we can take the KS determinant. Even in systems where the one-determinant approximation is not particularly good, such as small molecules, the overlaps of the Dyson orbitals of the primary ion states with the occupied orbitals are very close to 1.0 (*ca.* 0.99, see ref. 25). It is consistent with the unchanging nature of the KS orbitals from C^- to C that we have $\bar{d}_i = d_i = \psi_i$. Setting in eqn (24) both $\bar{I}_i = I_i$ and $\bar{d}_i = d_i$ we obtain

$$\Delta v^{N-1} = (I - A) \sum_i \frac{|d_i(1)|^2}{\rho(1)} = (I - A) \text{ for } r_1 \in \text{crystal} \quad (26)$$

So we conclude from (23) and (26) that

$$\Delta v^{\text{resp}} = v^{\text{resp}}[C^-] - v^{\text{resp}}[C] = \Delta v^{N-1} - \Delta v_s^{N-1} = \Delta \quad (27)$$

We have earlier given physical arguments (based on asymptotic decay of the density and the highest occupied orbital) that the KS potential must shift up by a constant going from C to C^- . We now see that this constant upshift is consistent with an exact expression for the KS potential, which does indeed lead to such a constant upshift. It is the response part of the KS potential that provides the upshift.

IV. Calculation of the difference Δ between KS band gap and fundamental gap

Bandstructure calculations yield the KS band gap. It is not possible to calculate the positive and negative ions of macroscopically sized crystals in order to obtain the fundamental gap. Fortunately, the special properties of crystals which are a consequence of their large size, make it easily possible to calculate the correction Δ to the KS band gap so as to obtain the fundamental gap to the same accuracy as is afforded by the standard density functional approximations for many other properties. It is obvious from the above that a good approximation to the response potential is required. We will demonstrate that the approximation of v^{resp} proposed in ref. 30 together with the simplifications that arise from the macroscopic dimensions of crystals afford accurate calculations of Δ and the fundamental gap ($I - A$) once the band structure and its KS gap are available.

We have already noted the step behavior of the v^{N-1} and v_s^{N-1} potentials (and hence the response potential) in atoms. This agrees with the derivation of the exchange part of the response potential by Krieger, Li and Iafrate^{27,62}

$$v_{\text{xresp}}^{\text{KLI}}(r_1) = \sum_{i=1}^H w_i \frac{|\psi_i(r_1)|^2}{\rho(r_1)} \quad (28)$$

where the KLI constant w_i is zero for the HOMO spinorbitals, $w_i = 0$, $i = H\alpha, H\beta$. In ref. 30 the form

$$v_{\text{xresp}}^{\text{GLLB}}(r_1) = \sum_{i=1}^H w_i^{\text{GLLB}} \frac{|\psi_i(r_1)|^2}{\rho(r_1)} \quad (29)$$

was proposed where the step height w_i^{GLLB} is estimated with the square root of the orbital energy difference

$$w_i^{\text{GLLB}} = K \sqrt{\varepsilon_H - \varepsilon_i} \quad (30)$$

(K is a constant). The square root behavior of the model (30) was observed empirically in atomic calculations.³⁰ Moreover it provides the proper scaling property of an exchange potential

$$v_{\text{xresp}}^{\text{GLLB}}([\rho_\lambda]; r_1) = \lambda v_{\text{xresp}}^{\text{GLLB}}([\rho]; \lambda r_1), \quad (31)$$

where

$$\rho_\lambda(r_1) = \lambda^3 \rho(\lambda r_1). \quad (32)$$

For solid state calculations it is important that the form (29) and (30) can be shown to be the exact exchange response potential in the limit of the homogeneous electron gas.³⁰ From the electron gas the value $K = 0.382$ is obtained. The response potential $v_{\text{xresp}}^{\text{GLLB}}$, however, is not just correct in the electron gas, it essentially restores the orbital dependence that gives the response potential its proper structure and strength in finite systems such as atoms and molecules. It has recently been demonstrated that this response potential is in molecules far better than the LDA one and GGA ones. It leads to a large improvement in the orbital energy spectrum, reducing by an order of magnitude the discrepancy between the HOMO orbital energy and $-I$, which is 4–6 eV with LDA and GGA. We note that $v_{\text{xresp}}^{\text{GLLB}}$ of (29) correctly possesses the zero last step ($w_H^{\text{GLLB}} = 0$) which is characteristic for the exact response potential, see discussion of eqn (10) and (17)–(19), and which has also been obtained in the KLI approximation.^{27,62} Adding the xc hole potential the GLLB approximation to the exchange–correlation potential now becomes

$$v_{\text{xc}}^{\text{GLLB}} = v_{\text{xc}}^{\text{hole}} + \sum_i^H K_{\text{xc}} \sqrt{\varepsilon_H - \varepsilon_i} \frac{|\psi_i|^2}{\rho} \quad (33)$$

We have noted that the exchange–correlation hole potential is twice the exchange–correlation density (7), which makes it possible to obtain approximations for the hole potential from approximate xc energy functionals. Of course $\varepsilon_{\text{xc}}(1)$ is not a unique function of position, but most approximate xc functionals do use an approximation for the hole potential to approximate $\varepsilon_{\text{xc}}(1)$. This is particularly true for exchange-only LDA, where $v_{\text{xhole}}^{\text{LDA}}(1) = 2\varepsilon_{\text{x}}^{\text{LDA}}(1)$. It is also possible to incorporate the correlation energy density, and possibly introduce modifications that introduce proper asymptotic $-1/r$ behavior of the energy density and potential, for instance as in ref. 63: $v_{\text{xhole}}(1) \approx 2\varepsilon_{\text{xc}}^{\text{LDAmod}}[\rho](1)$. Kuusma *et al.*³¹ have used the hole potential from the energy density of the PBEsol functional, $v_{\text{xhole}}^{\text{GLLB-sc}}(1) = 2\varepsilon_{\text{xc}}^{\text{PBEsol}}[\rho](1)$. These choices have some effect on the SCF calculations (band structure calculations) but they do not affect our derivation of a method for calculating the shift Δ , as will be clear below.

We need to calculate the quantity

$$\begin{aligned} \Delta &= \varepsilon_L(N+1) - \varepsilon_L(N) \equiv \bar{\varepsilon}_L - \varepsilon_L \\ &= \langle \bar{\psi}_L | \hat{t} + v_{\text{ext}} + \bar{v}_H[\bar{\rho}] + \bar{v}_{\text{hole}}[\bar{\rho}] | \bar{\psi}_L \rangle \\ &\quad + \left\langle \bar{\psi}_L \left| \sum_i^L K_{\text{xc}} \sqrt{\varepsilon_L - \varepsilon_i} \frac{|\bar{\psi}_i(N+1)|^2}{\bar{\rho}} \right| \bar{\psi}_L \right\rangle \\ &\quad - \langle \psi_L | \hat{t} + v_{\text{ext}} + v_H[\rho] + v_{\text{hole}}[\rho] | \psi_L \rangle \\ &\quad - \left\langle \psi_L \left| \sum_i^H K_{\text{xc}} \sqrt{\varepsilon_H - \varepsilon_i} \frac{|\psi_i(N)|^2}{\rho} \right| \psi_L \right\rangle \end{aligned} \quad (34)$$

The quantities of the negative ion (the $(N+1)$ -electron system) are indicated with an overbar. However, we have already noted (see Fig. 3) that due to the large dimensions of our system, over the bulk of the crystal (where the expectation

values of eqn (34) are made up) the density is unchanged, $\bar{\rho}(1) = \rho(1)$, and the Hartree and hole potentials are unchanged as well, $\bar{v}_H[\bar{\rho}] = v_H[\rho]$ and $\bar{v}_{\text{hole}}[\bar{\rho}] = v_{\text{hole}}[\rho]$. Since also the LUMO ψ_L will be unchanged from the (N) - to the $(N + 1)$ -electron system, the contributions from the first four operators cancel. Only the contributions from the response potentials play a role, in agreement with Sections IIIB and C. These do not cancel. We have already observed that the orbital energies $\bar{\epsilon}_i$ of the $(N + 1)$ system are shifted up by Δ , so all $(\bar{\epsilon}_L - \bar{\epsilon}_i)$ will be equal to the neutral crystal $(\epsilon_L - \epsilon_i)$. Substituting into eqn (34) we obtain the simple expression

$$\Delta = \left\langle \psi_L \left| \sum_i^H K_{xc} [\sqrt{\epsilon_L - \epsilon_i} - \sqrt{\epsilon_H - \epsilon_i}] \frac{|\psi_i|^2}{\rho} \right| \psi_L \right\rangle \quad (35)$$

where all quantities come from the ground state band structure calculation. In principle the integration runs over those spatial regions where the orbital ψ_L has sufficient amplitude that significant contributions to the normalization integral are obtained. This is the entire crystal but not the asymptotic regions. The integration can be replaced by integration over a unit cell, where normalization over the unit cell is applied.

The GLLB response potential is an approximation to the exact response potential. This is evident from the fact that it is not a constant over the crystal. We may nevertheless test how well this approximation is doing by calculating Δ with eqn (35) and comparing the resulting fundamental gap to experiment.

Kuisma *et al.*³¹ have been the first to recognize that this efficient calculation of the correction Δ is possible with the GLLB response potential, and we first review some of their results.

In Fig. 4 the KS one-electron energy gap from self-consistent field calculations with the GLLB-sc one-electron Hamiltonian are plotted as a function of the lattice constant (broken lines) in Si and C (left panel) and in AlAs, GaAs, and Ge (right panel). Over the variation by *ca.* 2% of the lattice constant, the variation in the KS gap is moderate (*ca.* 0.1 eV) for Si and C, but more than 0.5 eV for Ge. The KS gaps with the LDA, PBE and PBEsol functionals (each one only indicated at its optimized lattice constant) are somewhat smaller than the GLLB-sc ones. As will be clear presently (see discussion of Fig. 5) the GLLB-sc KS gaps agree better with benchmark EXX – RPA values for the KS gap. The fundamental gaps, indicated with an asterisk at the experimental lattice constant, are considerably higher than all approximate KS gaps, including the GLLB-sc one, as expected. The GLLB-sc fundamental gaps as a function of lattice constant (drawn lines) are obtained by correcting the GLLB-sc KS gaps with the constant Δ evaluated with eqn (35) after the band structure calculation. For Si, C and AlAs the experimental fundamental gap is very well reproduced. For Ge and GaAs the theoretical fundamental gap is still too small. From Fig. 5 it is clear that the error is not in the GLLB-sc KS gaps, which for these compounds are larger and much better than the LDA and GGA (PBE and PBEsol) KS gaps. However, the correction Δ is too small. It is to be noted that the correction for GaAs is much smaller than for Si, while the KS gap is even larger. If just the

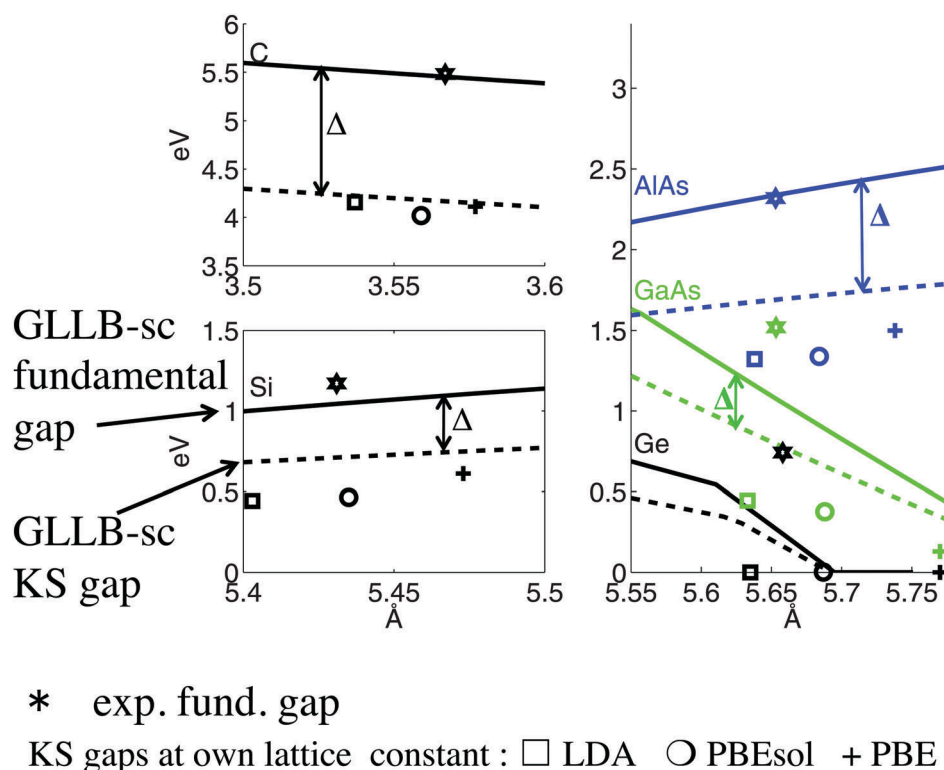


Fig. 4 KS gaps (broken lines) from self-consistent GLLB-sc calculations as a function of the lattice constant. Fundamental gaps as a function of the lattice constant (drawn lines) from the GLLB response potential correction (Δ) to the KS gaps. KS gaps from PBE (+), PBEsol (○) and LDA (□) self-consistent calculations are each given at its own lattice constant. Experimental fundamental gap (★) is given at the experimental lattice constant. Reproduced with permission from M. Kuisma, J. Ojanen, J. Enkovaara and T. T. Rantala, *Phys. Rev. B: Condens. Matter Mater. Phys.* 2010, **82**, 115106.³¹

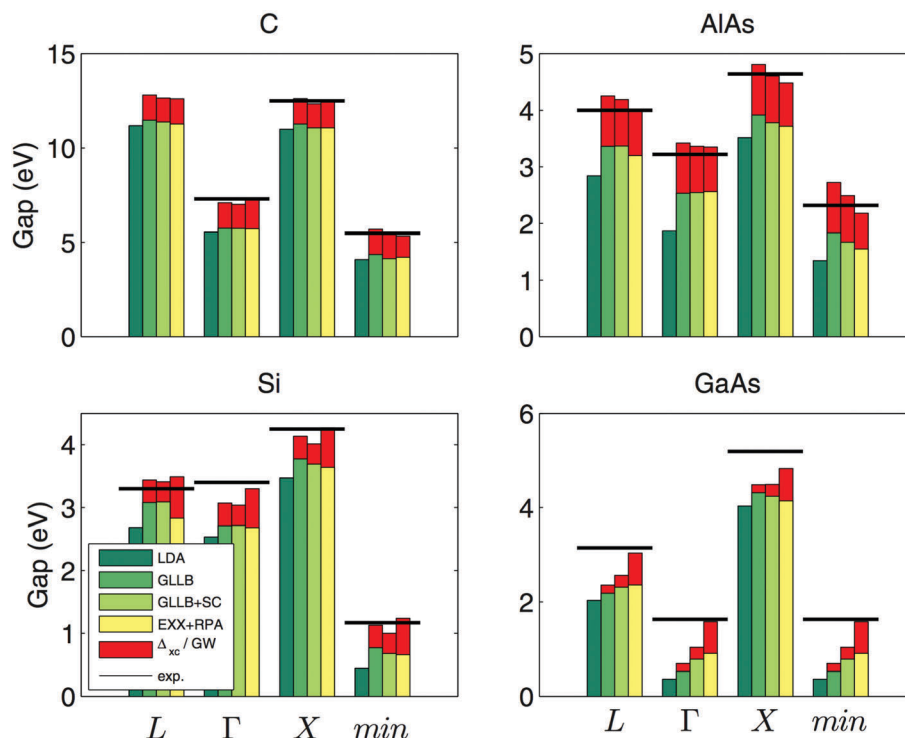


Fig. 5 KS gaps at different \mathbf{k} points from various KS one-electron energy calculations: LDA (dark green), GLLB-exchange-only (mid-green), GLLB-sc (light green). The comparison is to the accurate KS gaps from EXX – RPA calculations (yellow). Δ (red) from the GLLB response potential added in each case to the direct gap. The Δ (red) from GW calculations is added to the EXX + RPA KS gap to obtain a benchmark quasiparticle gap. Experimental estimates are shown as a bar on top. Reproduced with permission from M. Kuusma, J. Ojanen, J. Enkovaara and T. T. Rantala, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **82**, 115106.³¹

$i = \text{H}$ term in (35) would make the largest contribution, then a result proportional to $\sqrt{\varepsilon_L - \varepsilon_H} = \sqrt{g^{\text{KS}}}$ would be expected. Note that Ge illustrates that when the KS gap is zero, $g^{\text{KS}} = \varepsilon_L - \varepsilon_H = 0$, then the correction according to eqn (35) becomes zero too (if $\varepsilon_L = \varepsilon_H$ all contributions are zero). Quite generally, in metals Δ is zero, as it should be. However, if a system is erroneously calculated to be a metal in some density functional approximation, the present method does not provide a correction. For Ge at lattice constant 5.69 Å and larger $g^{\text{KS}} = 0$, *i.e.* Ge becomes metallic with GLLB-sc and there is no Δ correction from our formula.

In Fig. 5 the KS gaps are given at a few distinct \mathbf{k} points (direct gaps), as well as the corrections Δ (red) calculated with the lowest unoccupied Bloch states at those \mathbf{k} points. Benchmark values, *i.e.* EXX + RPA KS gaps (yellow), and Δ corrections on top of those values (red) as well as experimental fundamental gaps (black lines) are taken from ref. 41. As has been observed before,⁴¹ the LDA gaps (dark green) are quite close to the “true” KS gaps (EXX + RPA, yellow), although they definitely tend to be too low. The GLLB-exchange-only (mid-green) and GLLB-sc (light green) KS gaps correct for this error and are close to the benchmark EXX + RPA KS gaps (yellow). It is to be noted that the minimum gap is not at a specific \mathbf{k} point (a direct gap), the top of the valence band and the bottom of the conduction band occurring at different \mathbf{k} points (except for GaAs where it is the direct gap at the Γ point). The minimum gap is considerably smaller than the direct gaps at the given high-symmetry \mathbf{k} points, and the Δ contribution calculated at the \mathbf{k} point of the bottom of the conduction band (very similar of course to Δ at

other \mathbf{k} points) constitutes a considerable percentage of the minimal gap (the fundamental gap). After addition of Δ from our approximate v^{resp} to the GLLB-sc KS gaps, the calculated fundamental gaps have similar agreement with experiment as the best available estimates of the quasiparticle gap from EXX – RPA KS gaps with Δ correction from GW. Only for GaAs the latter are in better agreement with experiment.

The group of Thygesen and Jacobsen and coworkers have carried out many tests and comparisons with the GLLB-sc functional.^{32–37} We quote from ref. 34 a test for a series of semiconductors, see Fig. 6. The figure shows the accurate correction of the heavily underestimated LDA gaps by the GLLB-sc calculation (including the correction Δ). Many tests have included 2D materials, mostly transition metal dichalcogenides and oxides, *e.g.* the database of 51 of these materials in ref. 36, see Fig. 7, and for instance 27 organometallic halide perovskites ($\text{C}_4\text{H}_9\text{NH}_3)_2\text{MX}_2\text{Y}_2$, ($\text{M} = \text{Pb}, \text{Ge}, \text{Sn}$ and $\text{X}, \text{Y} = \text{Cl}, \text{Br}, \text{I}$), and derived structures with different types of vacancies.³⁷ The calculations are relatively fast, much cheaper than hybrid calculations (little effort above that of the LDA calculation), so extensive screening of potentially efficient semiconductors for the absorption of solar radiation could be performed (*e.g.* 5400 materials in ref. 32).

V. Conclusions

We summarize the essential differences between the KS gaps and fundamental gaps in molecules (*cf.* ref. 6 and 7) and solids.

Bandgaps of Si, InP, GaAs, AlP, ZnO, ZnS, C, BN, MgO, LiF

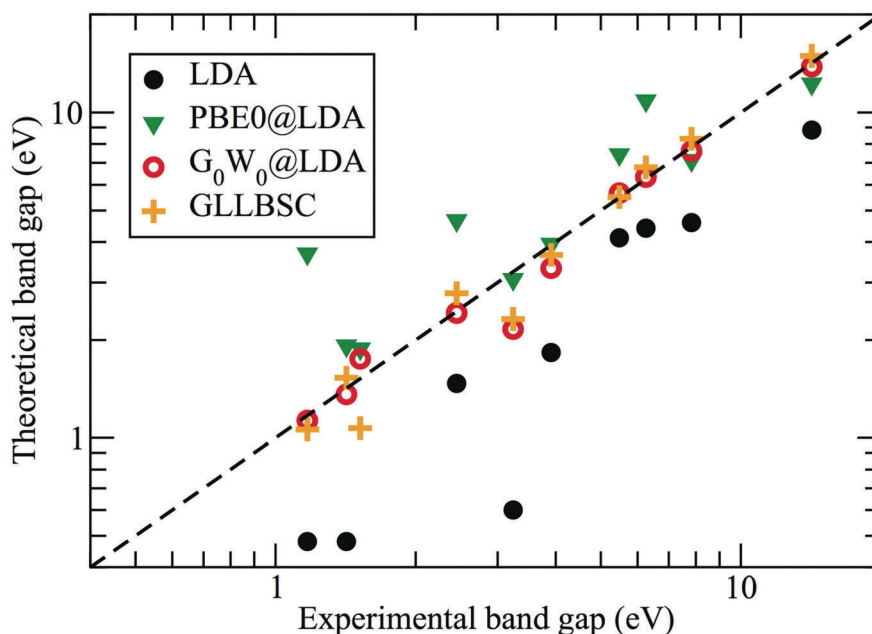


Fig. 6 Comparison of calculated³⁴ and experimental band gaps of Si, InP, AlP, ZnO, ZnS, C, BN, MgO, LiF. G_0W_0 @LDA refers to fully frequency dependent non-SCF GW based on LDA. The PBE0 results are obtained nonselfconsistently using LDA orbitals. Note the logarithmic scale; the LDA gap for GaAs (exp. band gap 1.52 eV) is 0.38 eV, too low to be visible on the figure. Reproduced with permission from F. Hüsler, T. Olsen and K. S. Thygesen, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 235132.³⁴

[M stands for neutral molecule, M^- for its negative ion; C stands for the neutral crystal, C^- for the crystal with one electron added. H is the N -th spinorbital (the HOMO in the neutral systems); L is the $(N + 1)$ -th spinorbital, the LUMO in the neutral systems and the HOMO in the negative ions.]

• Molecules:

Experiment: the optical gap is much smaller than the fundamental gap ($I - A$);

KS system: $\varepsilon_H[M] = -I$; $\varepsilon_L[M] \ll -A$; $\varepsilon_L[M^-] = -A$;

The KS gap $\varepsilon_L[M] - \varepsilon_H[M] \approx$ optical gap \ll fundamental gap ($I - A$)

• Solids:

Experiment: gap to delocalized excited states \approx fundamental gap ($I - A$);

KS system: $\varepsilon_H[C] = -I$; $\varepsilon_L[C] < -A$; $\varepsilon_L[C^-] = -A$

The KS gap $\varepsilon_L[C] - \varepsilon_H[C] <$ (optical) delocalized excited states gap \approx fundamental gap ($I - A$)

From the results of this paper we conclude that the fundamental gap in solids can be calculated straightforwardly, after a band structure calculation. The correction Δ to be applied to the KS or DFA band gap $\varepsilon_L - \varepsilon_H$ to obtain the fundamental gap $I - A$ can be obtained from the response part of the exchange–correlation potential v_{xc} , for which the GLLB approximation seems to be adequate. In the homogeneous electron gas (EG) GLLB is not an approximation but is exactly the exchange part of the response potential. This does not mean that the present

GLLB approximation of Δ is a local-density approximation. It is an essentially nonlocal contribution, based on orbitals and orbital energies. Ref. 46 has made the point that in the limit of infinitely extended systems simple functionals like LDA cannot provide the correction Δ to the band gap, not even with a proper ensemble definition of the energy terms.

We have identified the discrepancy between KS band gap and fundamental gap (*i.e.* the deviation of ε_L from $-A$, the “band gap problem”) as being due to the difference in the response part of the KS potential between the negative ion (crystal with electron added) and the neutral crystal. This also explains some of the confusion that has surrounded this topic from the beginning. It has initially been thought that on simple physical grounds⁵⁵ the level ε_L should be at $-A$. Indeed, the extended nature of solids causes the physical terms in the KS potential, such as the xc hole potential and the Hartree potential, to be identical in the neutral crystal C and the ion C^- . This suggests that ε_L would not change going to C^- . Being the orbital energy of the highest occupied KS level of C^- , it should be equal to minus the ionization energy of C^- , *i.e.* $-A$, and this would then already be the case in the neutral C system. However, this argument fails on two counts. First, the KS potential also has components, like the response potential, that cannot be understood so readily from elementary physical considerations based on Coulombic potentials of the charge density and the exchange–correlation hole. The KS potential is

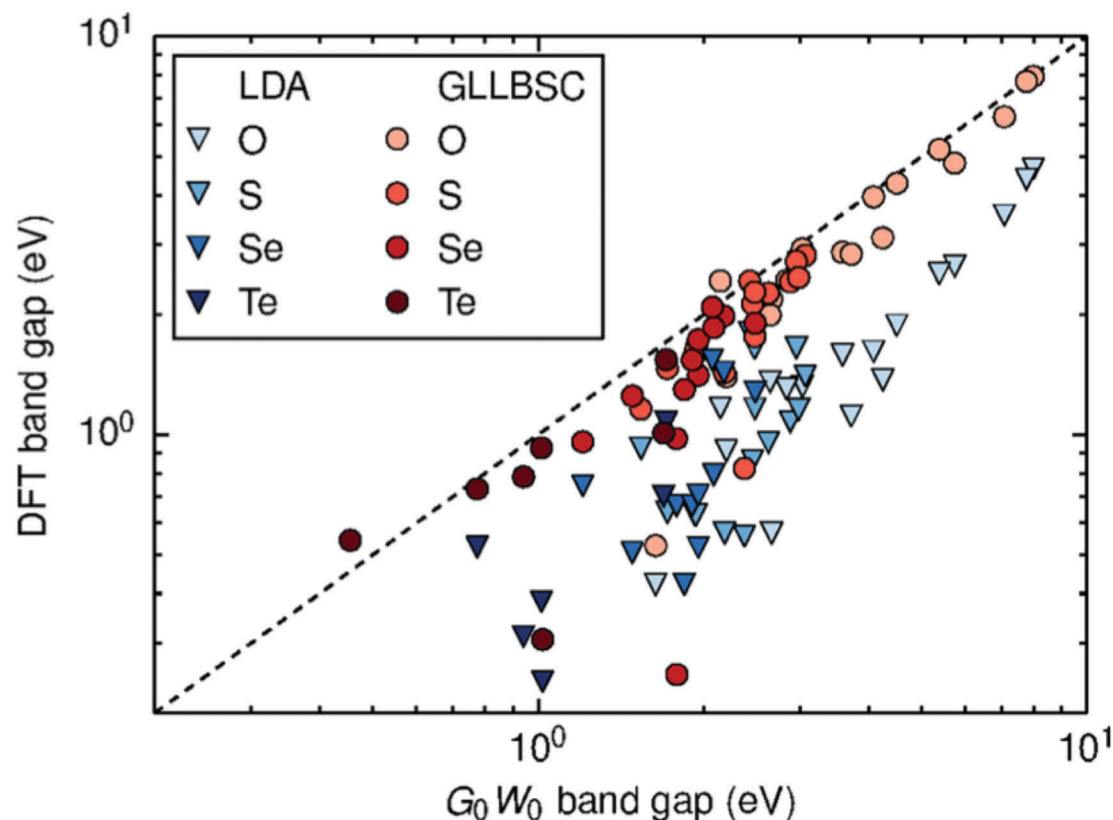


Fig. 7 Computed band gaps from ref. 36 of 51 monolayer TM dichalcogenides MX_2 ($\text{X} = \text{S}, \text{Se}, \text{Te}$) and TM oxides. The G_0W_0 approximate fundamental gaps are compared to the KS band gaps from LDA calculations at the PBE optimized structures and to fundamental gaps (including the Δ correction) from the GLLB-sc functional. Note the logarithmic scale. Reproduced with permission from F. A. Rasmussen and K. S. Thygesen, *J. Phys. Chem. C*, 2015, **119**, 13169.

mathematically defined as that local potential that reproduces the exact density. It may exhibit features that are required to achieve this goal, but that are not so easily interpretable in simple physical terms. The response potential is obviously a part of the KS potential that cannot be traced to some simple Coulombic potential of an electron density or hole density. In the second place we have stressed that the presence of the potential of the rather localized xc hole means that the KS potential has an attractive component that is not mirrored in a stabilizing hole potential for the electron-hole pair created in a completely delocalized $\text{H} \rightarrow \text{L}$ excitation. A delocalized excitation has a completely delocalized hole, that has no stabilizing effect. Such stabilization by a hole potential is also lacking for an added electron. The presence of the xc hole potential invalidates the expectation that ε_{L} will be at $-A$.

It is fortunate that the effect of the response potential, which we have identified as the source of the upshift Δ , can be approximated easily and rather accurately with the available approximate expression of ref. 30. This method is substantially cheaper than existing successful methods for computing fundamental gaps with hybrid functionals.^{10,11}

We have only considered the integer electron systems C and C^- . It has not been necessary to invoke the theory of Perdew, Parr, Levy and Balduz (PPLB)⁴³ of fractional electron systems (defined by ensembles of integer electron systems). This theory also introduces a uniform upshift of the potential, by a

constant of the same magnitude $-A - \varepsilon_{\text{L}}$, in a fractional electron system as soon as the electron number surpasses the integer N . But the “band gap problem” in solids can be understood and be solved in good approximation without any reference to the special case of fractional electron systems or ensembles. Δ is often called the derivative discontinuity, maybe as a reminder that in the PPLB theory not only does the KS potential shift up by Δ at $N + \delta$ electrons, but also the derivative of the total energy has a discontinuity at N , changing from $-I$ for $N - \delta$ to $-A$ for $N + \delta$. It is confusing to call Δ in the context of the band gap problem “the derivative discontinuity”, since the discrepancy between ε_{L} and $-A$ has nothing to do with fractional electron systems (ensembles). It is true that $\Delta = -A - \varepsilon_{\text{L}}$ also features in the PPLB theory of fractional electron systems, but it is an inversion of logic to then invoke the fractional electron system as “explaining” or even “being the cause” of the KS band gap/fundamental gap discrepancy in solids.

Appendix A: derivation of the response potential

The response potential can be derived¹² from the Schrödinger equation for the square root of the density.^{64–66} Applying this equation both in the case of the interacting electron system and

in the case of the KS system of noninteracting electrons, and using the equality of these densities, we can derive the useful expression for the response potential used in this paper.^{22,25,67}

Let us define the conditional amplitude, both in the case of the interacting electron system in its ground state Ψ_0^N and for the noninteracting KS system in its ground state $\Psi_{s,0}^N$,

$$\Phi(2 \dots N|1) = \frac{\Psi_0^N}{\sqrt{\rho(1)/N}}; \quad \Phi_s(2 \dots N|1) = \frac{\Psi_{s,0}^N}{\sqrt{\rho(1)/N}} \quad (\text{A1})$$

The conditional amplitude is a normalized $(N - 1)$ -electron wavefunction, the square of which gives the probability to find electrons at positions $2 \dots N$ when one electron is known to be at 1 (at position \mathbf{r}_1 with spin s_1). The conditional density $\rho^{\text{cond}}(2|1)$ is defined as the probability to find an electron at 2 when an electron is known to be at 1, and can be seen to be the electron density of the normalized $(N - 1)$ -electron wavefunction $\Phi(2 \dots N|1)$ ($\Gamma(1,2)$ is the two-particle density matrix),

$$\Gamma(1,2) = N(N-1) \int \Psi_0^{N*} \Psi_0^N d3 \dots dN \quad (\text{A2})$$

$$\begin{aligned} \rho^{\text{cond}}(2|1) &= \frac{\Gamma(1,2)}{\rho(1)} = \frac{N(N-1)}{\rho(1)} \int \Psi_0^{N*} \Psi_0^N d3 \dots dN \\ &= (N-1) \int \Phi^*(2 \dots N|1) \Phi(2 \dots N|1) d3 \dots dN \end{aligned} \quad (\text{A3})$$

Since the two-particle density matrix can be broken up in the uncorrelated probability to find two particles at positions 1 and 2, $\rho(1)\rho(2)$, plus the modification brought about by exchange and correlation effects, $\Gamma_{\text{xc}}(1,2)$, we see that $\rho^{\text{cond}}(2|1)$ is the full density $\rho(2)$ plus the exchange–correlation hole surrounding the electron at position 1 (it defines the exchange–correlation hole):

$$\begin{aligned} \rho^{\text{cond}}(2|1) &= \rho(2) + \frac{\Gamma_{\text{xc}}(1,2)}{\rho(1)} \\ &\equiv \rho(2) + \rho_{\text{xc}}^{\text{hole}}(2|1) \\ v^{\text{cond}}(1) &= v_{\text{H}}(1) + v_{\text{xc}}^{\text{hole}}(1) \end{aligned} \quad (\text{A4})$$

Similar definitions can be given for the KS system, where the full density is the same and the hole only consists of the exchange hole.

We now partition the N -electron Hamiltonian into an $(N - 1)$ -electron Hamiltonian for electrons $2 \dots N$, and project the Schrödinger equation against the conditional amplitude, writing $\sqrt{\rho(1)/N} = f(1)$

$$\begin{aligned} \hat{H}^N(1 \dots N) &= -\frac{1}{2} \nabla^2(1) + v_{\text{nuc}}(1) \\ &+ \sum_{p>1} \frac{1}{r_{1p}} + \hat{H}^{N-1}(2 \dots N) \end{aligned} \quad (\text{A5})$$

$$\begin{aligned} \hat{H}^N \Psi_0^N &= E_0^N \Psi_0^N \\ \int \Phi^* \hat{H}^N \Psi_0^N d2 \dots dN - E_0^{N-1} f(1) &= (E_0^N - E_0^{N-1}) f(1) \end{aligned} \quad (\text{A6})$$

Using the definition (A7) for the kinetic potential v_{kin} and definition (A8) for the potential v^{N-1} we obtain

$$v_{\text{kin}}(1) \equiv \int \Phi^*(2 \dots N|1) \left(-\frac{1}{2} \nabla^2(1) \right) \Phi(2 \dots N|1) d2 \dots dN \quad (\text{A7})$$

$$v^{N-1} \equiv \int \Phi^*(2 \dots N|1) \hat{H}^{N-1} \Phi(2 \dots N|1) d2 \dots dN - E_0^{N-1} \quad (\text{A8})$$

$$\begin{aligned} &\left[-\frac{1}{2} \nabla^2(1) + v_{\text{kin}}(1) + v_{\text{ext}}(1) + v^{\text{cond}}(1) + v^{N-1}(1) \right] f(1) \\ &= (E_0^N - E_0^{N-1}) f(1) = (-I) f(1) \end{aligned} \quad (\text{A9})$$

($v^{\text{cond}}(1)$ is the Coulomb potential at position 1 of the density $\rho^{\text{cond}}(2|1)$.)

We can also derive an equation for $f(1) = \sqrt{\rho(1)/N}$ from the KS model. The KS noninteracting electrons have the same one-electron density as the exact wavefunction. They move in the “external” potential $v_{\text{s}}(1)$ and do not have electron interaction so the Hamiltonian is

$$\hat{H}_{\text{s}}^N = \sum_{i=1}^N \left(-\frac{1}{2} \nabla^2(i) + v_{\text{s}}(i) \right) \quad (\text{A10})$$

Defining $v_{\text{s,kin}}(1)$ and v_{s}^{N-1} for the KS system in the same way as for interacting electrons,

$$v_{\text{s,kin}}(1) \equiv \int \Phi_{\text{s}}^*(2 \dots N|1) \left(-\frac{1}{2} \nabla^2(1) \right) \Phi_{\text{s}}(2 \dots N|1) d2 \dots dN \quad (\text{A11})$$

$$v_{\text{s}}^{N-1} \equiv \int \Phi_{\text{s}}^*(2 \dots N|1) \hat{H}_{\text{s}}^{N-1} \Phi_{\text{s}}(2 \dots N|1) d2 \dots dN - E_{\text{s},0}^{N-1} \quad (\text{A12})$$

we can write the Schrödinger equation for $\sqrt{\rho(1)/N}$ also as (note the absence of a conditional potential because of the zero electron–electron interaction)

$$\begin{aligned} &\left[-\frac{1}{2} \nabla^2(1) + v_{\text{s,kin}}(1) + v_{\text{s}}(1) + v_{\text{s}}^{N-1}(1) \right] f(1) \\ &= (E_{\text{s},0}^N - E_{\text{s},0}^{N-1}) f(1) = \varepsilon_N f(1) = (-I) f(1) \end{aligned} \quad (\text{A13})$$

(The total energy for the noninteracting electron system is just the sum of orbital energies.) From the density eqn (A13) and (A9) we arrive at the following expression for v_{s} and $v_{\text{xc}}^{12,22,23,25,68}$

$$\begin{aligned} v_{\text{s}}(1) &= v_{\text{ext}}(1) + v^{\text{cond}}(1) + v_{\text{kin}}(1) - v_{\text{s,kin}}(1) + v^{N-1} - v_{\text{s}}^{N-1} \\ v_{\text{c,kin}} &= v_{\text{kin}} - v_{\text{s,kin}}; \quad v^{\text{resp}} = v^{N-1} - v_{\text{s}}^{N-1} \\ v_{\text{s}}(1) &= v_{\text{ext}}(1) + v_{\text{H}}(1) + v_{\text{xc}}^{\text{hole}}(1) + v_{\text{c,kin}}(1) + v^{\text{resp}} \\ &= v_{\text{ext}}(1) + v_{\text{H}}(1) + v_{\text{xc}}(1) \end{aligned} \quad (\text{A14})$$

The name “response potential” is used since this part of the exchange–correlation potential can be associated with the response of the pair-correlation factor under density variation,

see eqn (5) in the Introduction (here we do not use the integration over the coupling constant). But the present expression for v^{resp} is more useful, since it allows us to express v^{resp} in an expansion over orbital contributions in the following way.

The potential v_s^{N-1} can be cast in a simple form employing the KS orbitals and orbital energies using the determinantal form of the KS wavefunction. Expand the determinant into the cofactors of the elements of the first column

$$\begin{aligned}\Psi_{s,0}^N(1, 2 \dots N) &= \frac{1}{\sqrt{N!}} \sum_{i=1}^N \psi_i^s(1) (-1)^{i+1} D_s^{(i1)}(2 \dots N) \\ &= \frac{1}{\sqrt{N}} \sum_i \psi_i^s(1) \Psi_s^{N-1}(\psi_i^{-1})\end{aligned}\quad (\text{A15})$$

where $D_s^{(i1)}$ denotes the (unnormalized) determinant with row i and column 1 crossed out. The $(1/\sqrt{(N-1)!})D_s^{(i1)}$ are normalized $(N-1)$ -electron ion wavefunctions of the noninteracting system resulting from ionization of the electron in orbital ψ_i^s , which we denote $\Psi_s^{N-1}(\psi_i^{-1})$. Upon substituting the expansion (A15) into the eqn (A12) for v_s^{N-1} one obtains

$$\begin{aligned}v_s^{N-1}(1) &= \frac{1}{\rho(1)} \int \sum_i \psi_i^s(1)^* \Psi_s^{N-1}(\psi_i^{-1})^* \hat{H}_s^{N-1}(2 \dots N) \\ &\quad \times \sum_j \psi_j^s(1) \Psi_s^{N-1}(\psi_j^{-1}) d2 \dots dN - E_{s,0}^{N-1} \quad (\text{A16}) \\ &= \frac{1}{\rho(1)} \sum_i |\psi_i^s(1)|^2 E_{s,i}^{N-1} - E_{s,0}^{N-1}.\end{aligned}$$

Since the total energies of the noninteracting states are just sums over the orbital energies, and using $\rho(1) = \sum_i |\psi_i^s(1)|^2$ it follows that

$$v_s^{N-1}(1) = \sum_i (\varepsilon_N - \varepsilon_i) \frac{|\psi_i^s(1)|^2}{\rho(1)} \quad (\text{A17})$$

The potential v_s^{N-1} has a remarkably simple form, which for instance in atoms leads to a clear step structure with step height $(\varepsilon_N - \varepsilon_i)$ in each shell where the density is almost completely built from orbitals of a shell, $i \in (nl)$. The step is zero in the region where the HOMO ψ_N^s (the most diffuse orbital) contributes exclusively to the density, *i.e.* in the outer region and asymptotically.

The potential v^{N-1} can be put into a similar form by using the Dyson expansion of the wavefunction in all ion wavefunctions (the set of all $(N-1)$ -electron wavefunctions $\{\Psi_i^{N-1}\}$),

multiplied with the corresponding Dyson orbitals d_i ,

$$\begin{aligned}d_i(1) &= \sqrt{N} \int \Psi_i^{N-1*} \Psi_0^N d2 \dots dN, \quad \rho(1) = \sum_i |d_i(1)|^2 \\ \sqrt{N} \Psi_0^N &= \sum_i d_i(1) \Psi_i^{N-1}(2 \dots N) \\ v^{N-1}(1) &= \sum_i (-I_0 + I_i) \frac{|d_i(1)|^2}{\rho(1)}\end{aligned}\quad (\text{A18})$$

where the expression for v^{N-1} in the last line follows again from substituting the expansion of Ψ_0^N of the second line into eqn (A8). Note that $i=0$ denotes the ion ground state Ψ_0^{N-1} , with Dyson orbital d_0 and ionization energy $I_0 = I$ (we write I_0 mostly as just I , the (first) ionization energy of the system). The Dyson orbital of the ion ground state is close to the HOMO (both to Hartree–Fock HOMO and the KS HOMO) and clearly the HOMO region (notably the asymptotic region) will have potential $v^{N-1} \rightarrow 0$. In the interacting system the Dyson orbitals of the primary ion states (those that are well approximated by a single determinant with a particular orbital crossed out) overlap strongly with orbitals of one-electron models like Hartree–Fock or Kohn–Sham. Satellites (small peaks in the photoelectron spectrum adjacent to the strong peaks corresponding to the primary (orbital) ionizations) correspond to ion states that in the orbital models have a leading determinant consisting of an orbital ionization plus an excitation. Their Dyson orbitals are similar in shape to the orbital characterizing the main peak but with much lower amplitude²⁵ (the Dyson orbitals are not normalized). Since the intensities of peaks in the photoelectron spectrum are proportional to the norms of the corresponding Dyson orbitals,⁶⁹ the main peak has high intensity, the satellites have low intensities.²⁵

The equation for v_s^{N-1} is obviously a simplification of the one for v^{N-1} : the occupied KS orbitals are just the Dyson orbitals of the KS noninteracting electron system for the primary ion states, *i.e.* the ion states characterized by a simple orbital ionization, $\Psi_s^{N-1}(\psi_i^{-1})$. There are of course many more ion states of the noninteracting system than just the primary ones, namely those consisting of a primary ionization (ψ_i^{-1}) plus a $\psi_j \rightarrow \psi_a$ (or double *etc.*) excitation, $\Psi_s^{N-1}(\psi_i^{-1}; \psi_j \rightarrow \psi_a)$ *etc.*²⁵ Those other ion states have Dyson orbitals that are identically zero. This follows from the determinantal nature of the ground state, of the primary ion wavefunctions, and of the “shake up” ion states, and from the simple rules for matrix elements of determinantal states,

$$\begin{aligned}d_s(\psi_i^{-1})(1) &= \sqrt{N} \int \Psi_s^{N-1}(\psi_i^{-1})^* \Psi_{s,0}^N d2 \dots dN = \psi_i^s(1) \\ d_s(\psi_i^{-1}; \psi_j \rightarrow \psi_a)(1) &= \sqrt{N} \int \Psi_s^{N-1}(\psi_i^{-1}; \psi_j \rightarrow \psi_a)^* \Psi_{s,0}^N d2 \dots dN \equiv 0\end{aligned}\quad (\text{A19})$$

There are no shake-up states (satellites in the photoemission spectrum) of a noninteracting electron system. This is why the “Dyson expansion” (A15) of the KS wavefunction has only N terms.

Acknowledgements

I thank Mikael Kuisma and Tapio Rantala for stimulating discussions at a PhD defense and beyond.

References

- 1 A. Görling, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1996, **54**, 3912.
- 2 A. Savin, C. J. Umrigar and X. Gonze, *Chem. Phys. Lett.*, 1998, **228**, 391.
- 3 J. Garza, J. Nichols and D. Dixon, *J. Chem. Phys.*, 2000, **113**, 6029.
- 4 M. Grüning, O. Gritsenko, S. J. A. van Gisbergen and E. J. Baerends, *J. Chem. Phys.*, 2002, **116**, 9591.
- 5 F. Della Sala, *Theor. Chem. Acc.*, 2007, **117**, 981.
- 6 E. J. Baerends, O. V. Gritsenko and R. van Meer, *Phys. Chem. Chem. Phys.*, 2013, **15**, 16408.
- 7 R. van Meer, O. V. Gritsenko and E. J. Baerends, *J. Chem. Theory Comput.*, 2014, **10**, 4432.
- 8 J. W. Kim, K. W. Hong, S. H. Choi, S. Y. Hwang and W. Y. Kim, *Phys. Chem. Chem. Phys.*, 2015, **17**, 31434.
- 9 O. V. Gritsenko, L. Mentel and E. J. Baerends, *J. Chem. Phys.*, 2016, **144**, 204114.
- 10 A. J. Garza and G. E. Scuseria, *J. Phys. Chem. Lett.*, 2016, **7**, 4165.
- 11 J. M. Crowley, J. Tahir-Kheli and W. A. Goddard III, *J. Phys. Chem. Lett.*, 2016, **7**, 1198.
- 12 M. A. Buijse, E. J. Baerends and J. G. Snijders, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1989, **40**, 4190.
- 13 E. J. Baerends and O. V. Gritsenko, *J. Phys. Chem. A*, 1997, **101**, 5383.
- 14 O. V. Gritsenko and E. J. Baerends, *Theor. Chem. Acc.*, 1997, **96**, 44.
- 15 R. van Leeuwen and E. J. Baerends, *Int. J. Quantum Chem.*, 1994, **52**, 711.
- 16 N. T. Maitra, *J. Chem. Phys.*, 2005, **122**, 234104.
- 17 N. Helbig, I. V. Tokatly and A. Rubio, *J. Chem. Phys.*, 2009, **131**, 224105.
- 18 D. G. Tempel, T. J. Martínez and N. T. Maitra, *J. Chem. Theory Comput.*, 2009, **5**, 770.
- 19 P. Elliott, J. I. Fuks, A. Rubio and N. T. Maitra, *Phys. Rev. Lett.*, 2012, **109**, 266404.
- 20 Z. J. Ying, V. Brosco, G. M. Lopez, D. Varsano, P. Gori-Giorgi and J. Lorenzana, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2016, **94**, 075154.
- 21 S. V. Kohut, A. M. Polgar and V. N. Staroverov, *Phys. Chem. Chem. Phys.*, 2016, **18**, 20938.
- 22 R. van Leeuwen, O. V. Gritsenko and E. J. Baerends, *Z. Phys. D: At., Mol. Clusters*, 1995, **33**, 229.
- 23 O. V. Gritsenko, R. V. Leeuwen and E. J. Baerends, *J. Chem. Phys.*, 1996, **104**, 8535.
- 24 E. J. Baerends and O. V. Gritsenko, *J. Phys. Chem. A*, 1997, **101**, 5383.
- 25 O. V. Gritsenko, B. Braïda and E. J. Baerends, *J. Chem. Phys.*, 2003, **119**, 1937.
- 26 E. J. Baerends and O. Gritsenko, *J. Chem. Phys.*, 2016, **145**, 037101.
- 27 J. B. Krieger, U. Li and G. J. Iafrate, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1992, **45**, 101.
- 28 O. Gritsenko, R. van Leeuwen and E. J. Baerends, *J. Chem. Phys.*, 1994, **101**, 8955.
- 29 D. P. Chong, O. V. Gritsenko and E. J. Baerends, *J. Chem. Phys.*, 2002, **116**, 1760.
- 30 O. V. Gritsenko, R. van Leeuwen, E. van Lenthe and E. J. Baerends, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1995, **51**, 1944.
- 31 M. Kuisma, J. Ojanen, J. Enkovaara and T. T. Rantala, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **82**, 115106.
- 32 I. E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen and K. W. Jacobsen, *Energy Environ. Sci.*, 2012, **5**, 5814.
- 33 J. Yan, K. W. Jacobsen and K. S. Thygesen, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **86**, 045208.
- 34 F. Hüser, T. Olsen and K. S. Thygesen, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 235132.
- 35 I. E. Castelli, F. Hüser, M. Pandey, H. Li, K. S. Thygesen, B. Seger, A. Jain, K. A. Persson, G. Ceder and K. W. Jacobsen, *Adv. Energy Mater.*, 2015, **5**, 1400915.
- 36 F. A. Rasmussen and K. S. Thygesen, *J. Phys. Chem. C*, 2015, **119**, 13169.
- 37 M. Pandey, K. W. Jacobsen and K. S. Thygesen, *J. Phys. Chem. Lett.*, 2016, **7**, 4346.
- 38 L. J. Sham and M. Schlüter, *Phys. Rev. Lett.*, 1983, **51**, 1888.
- 39 L. J. Sham and M. Schlüter, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1985, **32**, 3883.
- 40 R. W. Godby, M. Schlüter and L. J. Sham, *Phys. Rev. Lett.*, 1986, **56**, 2415.
- 41 R. W. Godby, M. Schlüter and L. J. Sham, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1987, **36**, 6497.
- 42 R. W. Godby, M. Schlüter and L. J. Sham, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 10159.
- 43 J. P. Perdew, R. G. Parr, M. Levy and J. Balduz, *Phys. Rev. Lett.*, 1982, **49**, 1691.
- 44 J. P. Perdew and M. Levy, *Phys. Rev. Lett.*, 1983, **51**, 1884.
- 45 A. J. Cohen, P. Mori-Sanchez and W. Yang, *Chem. Rev.*, 2012, **112**, 289.
- 46 E. Kraisler and L. Kronik, *J. Chem. Phys.*, 2014, **140**, 18A540.
- 47 M. Grüning, A. Marini and A. Rubio, *J. Chem. Phys.*, 2006, **124**, 154108.
- 48 J. Lim, S. H. Choi, J. W. Kim and W. Y. Kim, *J. Chem. Phys.*, 2016, **145**, 224309.
- 49 J. Rienstra-Kiracofe, G. S. Tschumper and H. F. Schaefer III, *Chem. Rev.*, 2002, **102**, 231.
- 50 R. van Leeuwen and E. J. Baerends, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1994, **49**, 2421.
- 51 O. Gritsenko and E. J. Baerends, *J. Chem. Phys.*, 2004, **121**, 655.
- 52 J. D. Chai and P. T. Chen, *Phys. Rev. Lett.*, 2013, **110**, 033002.
- 53 V. Brosco, Z. J. Ying and J. Lorenzana, *Sci. Rep.*, 2013, **3**, 2172.
- 54 J. F. Janak, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1978, **18**, 7165.

- 55 A. R. Williams and U. von Barth, in *Theory of the inhomogeneous Electron Gas*, ed. S. Lundquist and N. H. March, Plenum Press, New York, 1983.
- 56 M. A. Buijse and E. J. Baerends, in *Density functional theory of molecules, clusters and solids*, ed. D. E. Ellis, Kluwer Academic Publishers, The Netherlands, 1995, ch. 1, pp. 1–46.
- 57 M. Buijse and E. J. Baerends, *Mol. Phys.*, 2002, **100**, 401.
- 58 S. Fahy, X. W. Wang and S. G. Louie, *Phys. Rev. Lett.*, 1990, **65**, 1478, DOI: 10.1103/PhysRevLett.65.1478.
- 59 W. L. Luken and D. N. Beratan, *Theor. Chim. Acta*, 1982, **61**, 265.
- 60 J. Wang and E. J. Baerends, *J. Chem. Phys.*, 2013, **142**, 204311.
- 61 O. V. Gritsenko and E. J. Baerends, *J. Chem. Phys.*, 2004, **120**, 8364.
- 62 J. B. Krieger, U. Li and G. J. Iafrate, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1992, **46**, 5453.
- 63 O. V. Gritsenko, P. R. T. Schipper and E. J. Baerends, *Int. J. Quantum Chem.*, 2000, **76**, 407.
- 64 G. Hunter, *Int. J. Quantum Chem.*, 1975, **9**, 237.
- 65 G. Hunter, *Int. J. Quantum Chem., Symp.*, 1975, **9**, 311.
- 66 M. Levy, J. P. Perdew and V. Sahni, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1984, **30**, 2745.
- 67 D. P. Chong, O. V. Gritsenko and E. J. Baerends, *J. Chem. Phys.*, 2002, **116**, 1760.
- 68 O. V. Gritsenko and E. J. Baerends, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1996, **54**, 1957.
- 69 O. V. Gritsenko and E. J. Baerends, *Phys. Chem. Chem. Phys.*, 2016, **18**, 20945.