PHYSICAL REVIEW B **77**, 115123 (2008)

Fractional charge perspective on the band gap in density-functional theory

Aron J. Cohen, Paula Mori-Sánchez, and Weitao Yang
Department of Chemistry, Duke University, Durham, North Carolina 27708, USA

(Received 16 January 2008; published 18 March 2008)

The calculation of the band gap by density-functional theory (DFT) is examined by considering the behavior of the energy as a function of number of electrons. It is explained that the incorrect band-gap prediction with most approximate functionals originates mainly from errors in describing systems with fractional charges. Formulas for the energy derivatives with respect to number of electrons are derived, which clarify the role of optimized effective potentials in prediction of the band gap. Calculations with a recent functional that has much improved behavior for fractional charges give a good prediction of the energy gap and also $\varepsilon_{\mathrm{HOMO}} \simeq -I$ for finite systems. Our results indicate that it is possible, within DFT, to have a functional whose eigenvalues or derivatives accurately predict the band gap.

DOI: 10.1103/PhysRevB.77.115123 PACS number(s): 71.10.−w, 71.15.Mb

One of the many important uses of density-functional theory (DFT) is the calculation of the band structure which has many applications throughout physics, for example, in semiconductors, electron transport, and reactions at surfaces. The first step in achieving accuracy in the band structure is to understand the band gap which standard functionals have long been known to systematically underestimate by as much as ~50%. Recent efforts have focused on use of the optimized effective potential (OEP) method, which can often give an improvement in the prediction of band gaps for small-gap semiconductors, but has problems with wider gap semiconductors and insulators.[1–3] In many cases, it has proved necessary to move to the quasiparticle *GW* theory to calculate the band gap of solids accurately.[4] Conventionally, the band-gap problem has been related to the so-called "derivative discontinuity" in the exchange-correlation potential: even with an accurate Kohn-Sham potential, the energy gap is still different from the true gap by an amount of the derivative discontinuity.[5,6] This perspective, however, does not offer the understanding or the mechanism needed for making progress for band-gap prediction with DFT.

In this paper, a different perspective is offered: the band-gap problem is shown to be related to the behavior of approximate density functionals for fractional numbers of electrons, an issue which has drawn recent interest.[7–10] This enables us to understand the problem with band-gap calculations and offers ideas to develop functionals which predict the band gap correctly. Examples will be given for molecules where the energy gap can be compared with explicit calculations of systems with fractional charges, however, the same ideas apply to solids.

The fundamental band gap for an N_0-electron system in an external potential $v(\mathbf{r})$ is given by

$$E_{\mathrm{gap}}^{\mathrm{integer}} = [E_v(N_0-1) + E_v(N_0+1) - 2E_v(N_0)] = I - A, \tag{1}$$

where $E_v(N_0)$ is the ground-state energy of the N_0 particle system, I is the ionization energy, and A is the electron affinity. For a system with a fractional number of electrons $N_0+\delta N$, with $0<\delta N<1$, it has been shown that the energy is a straight line connecting the total energies at the integers; namely, $E_v(N_0+\delta N) = (1-\delta N)E_v(N_0) + \delta N E_v(N_0+1)$.[11,12] This linear relation means that the energy gap in Eq. (1) can be given by the derivative difference,

$$E_{\mathrm{gap}}^{\mathrm{deriv}} = \lim_{\delta N \to 0} \left\{ \left. \frac{\partial E_v}{\partial N} \right|_{N_0+\delta N} - \left. \frac{\partial E_v}{\partial N} \right|_{N_0-\delta N} \right\}. \tag{2}$$

If we substitute in the DFT the total energy expression $E_v = T_s[\rho] + V_{\mathrm{ext}}[\rho] + J[\rho] + E_{xc}[\rho]$, we have

$$E_{\mathrm{gap}}^{\mathrm{deriv}} = \varepsilon_{\mathrm{LUMO}}^{\mathrm{KS}} - \varepsilon_{\mathrm{HOMO}}^{\mathrm{KS}} + \Delta_{xc} = E_{\mathrm{gap}}^{\mathrm{KS}} + \Delta_{xc}, \tag{3}$$

where the $E_{\mathrm{gap}}^{\mathrm{KS}}$ is the gap in a Kohn-Sham calculation and the Δ_{xc} is the derivative discontinuity.[5,6]

In this work, we identify the problem with calculations using approximate density functionals by considering the basic assumption in Eq. (2), which the energy at N_0+1 and at N_0-1 can be given simply from the derivatives at N_0. This is true for exact DFT, but it may or may not be true for approximate functionals. The key is to investigate the behavior of the total energy as a function of numbers of electrons. To do this, we consider a noninteracting ground-state reference system where we allow the occupation numbers of the orbitals to vary the number of electrons smoothly; the first-order reduced density matrix of the reference system is given by

$$\rho_s(\mathbf{r},\mathbf{r}') = \sum_i n_i \phi_i(\mathbf{r}) \phi_i^*(\mathbf{r}'), \tag{4}$$

where $n_i=1$ for $i<f$, $n_i=\delta N$ for $i=f$, and $n_i=0$ for $i>f$, and f is the index for the frontier orbital. The behavior of three different exchange-correlation functionals: the local density approximation (LDA), Hartree-Fock (HF), and MCY3[13] are compared for self-consistent calculations[13] at fractional numbers of electrons in Fig. 1. MCY3 was constructed as a functional of $\rho_s(\mathbf{r},\mathbf{r}')$ to give a much improved description of fractional numbers of electrons, and it gives a straight line interpolation between the integers. The interpolation from the other functionals is far from linear, HF curves in a concave manner, and LDA in a very convex manner. There are two main problems with approximate density functionals. First, they can have a remarkably different behavior from the exact functional in fractional charge systems, compared to their behavior for the corresponding integer charge systems.

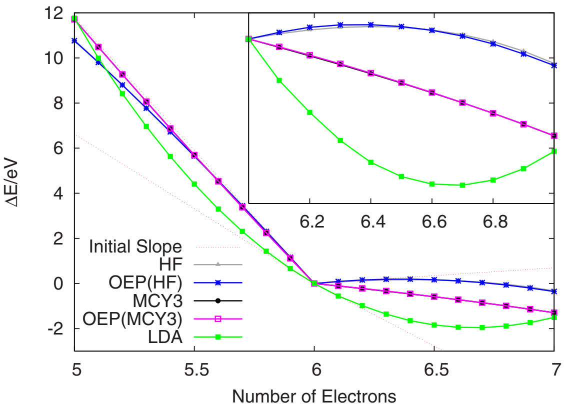

FIG. 1. (Color online) Difference energy of the carbon atom $[\Delta E=E(N_0+\delta N)-E(N_0)]$ against number of electrons $(N_0+\delta N, N_0=6)$ with several different functionals using OEP and GKS. Dotted line follows the initial slope for the nonstraight functionals. The inset shows $6<N<7$ range in more detail.

Second, the error in the integer charge systems can also be significant. The combined effects lead to the error in the band-gap prediction from derivative information in finite and infinite systems.

For molecules, LDA has a very reasonable description of the integer values (e.g., see Table 1 of Ref. 14) but a much worse description in between the integers. The use of the first derivatives for LDA will clearly not give the I and A from the integer calculations. Because of the convexity of the curve, LDA will give too small value for I and too large value for A, meaning that the band-gap $I-A$ will be too small, as shown by the dotted lines in Fig. 1. The case for HF is very different as the integer values are not as good because of the lack of correlation, and also it curves in a concave manner. These two errors cancel each other to some extent in the prediction of I but add together in the calculation of A. For HF, the value of I will be about right and A too small, meaning that the band-gap $I-A$ will be too large, as shown in Fig. 1. For MCY3, as it is very straight, the use of the derivatives will give a prediction very similar to the integer calculations for I and A, and the band-gap $I-A$ should be accurate as the integer values are good. For functionals that have a linear behavior for fractional charge systems (e.g., MCY3 and the exact functional), the initial derivative is all that is needed to calculate I and A and the band gap as in Eq. (2).

We now consider analytic expressions for $\partial E_v/\partial N$, the initial derivative, for some families of exchange-correlation functionals. The main ideas and equations are applicable to both molecules and solids and they are summarized here, with further details presented in the Appendix. In a noninteracting system with fractional charge [Eq. (4)], the orbitals are the eigenstates of a one-electron potential, either a local potential $v_s(\mathbf{r})$, in the original Kohn-Sham (KS) method,

$$\left(-\frac{1}{2}\nabla^2+v_s\right)|\phi_i^{\mathrm{KS}}\rangle=\varepsilon_i^{\mathrm{KS}}|\phi_i^{\mathrm{KS}}\rangle, \tag{5}$$

or a nonlocal potential $v_s^{\mathrm{NL}}(\mathbf{r},\mathbf{r}')$.

$$\left(-\frac{1}{2}\nabla^2+v_s^{\mathrm{NL}}\right)|\phi_i^{\mathrm{GKS}}\rangle=\varepsilon_i^{\mathrm{GKS}}|\phi_i^{\mathrm{GKS}}\rangle, \tag{6}$$

in the case of the generalized Kohn-Sham (GKS) method,[16] which has also been called the Hartree-Fock-Kohn-Sham method.[15] For a fixed minimizing potential v_s (either local or nonlocal), the energy is stationary with respect to the change in the potential; the corresponding orbitals and eigenvalues are therefore fixed and only the frontier level occupation n_f is able to change $\delta N=\delta n_f$. Thus, we have the key result,

$$\frac{\partial E_v}{\partial N}=\left(\frac{\partial E_v}{\partial n_f}\right)_{v_s}, \tag{7}$$

where the frontier orbital is either the lowest unoccupied molecular orbital (LUMO), $n_f=n_{\mathrm{LUMO}}$, if $\delta N>0$, or the highest occupied molecular orbital (HOMO), $n_f=n_{\mathrm{HOMO}}$, if $\delta N<0$.

We now consider three cases for which the analytic derivatives can be obtained.

Case A. $E_{xc}[\rho_s(\mathbf{r})]$, an explicit functional of ρ_s [e.g., LDA or generalized gradient approximation (GGA)],

$$\frac{\partial E_v}{\partial N}=\varepsilon_f^{\mathrm{KS}}, \tag{8}$$

where $\varepsilon_f^{\mathrm{KS}}$ is the KS eigenvalue for the frontier orbital in the local potential $v_s(\mathbf{r})=v(\mathbf{r})+v_J(\mathbf{r})+v_{xc}(\mathbf{r})$. This is exactly the combination of Eq. (7) with Janak's theorem for n_f.[17]

Case B. $E_{xc}[\rho_s(\mathbf{r},\mathbf{r}')]$, a functional of the first-order density matrix minimized with a local potential as in Eq. (5) [e.g., OEP(HF) and OEP(MCY3)],

$$\frac{\partial E_v}{\partial N}=\varepsilon_f^{\mathrm{KS}}+\langle\phi_f^{\mathrm{KS}}|v+v_J+v_{xc}^{\mathrm{NL}}-v_s|\phi_f^{\mathrm{KS}}\rangle=\varepsilon_f^{\mathrm{KS}}+\Delta_{xc}^f, \tag{9}$$

where the nonlocal potential $v_{xc}^{\mathrm{NL}}(\mathbf{r},\mathbf{r}')=\frac{\delta E_{xc}[\rho_s(\mathbf{r},\mathbf{r}')]}{\delta\rho_s(\mathbf{r},\mathbf{r}')}$. Equation (9) is a key result, showing that for general orbital functionals, $\partial E_v/\partial N$ is not given by the frontier OEP eigenvalue $\varepsilon_f^{\mathrm{KS}}$, but, with a correction term, the derivative discontinuity Δ_{xc}^f. This general result is related to the results of Ref. 1 in the case of exact exchange and of Ref. 18 from the self-energy.

Case C. $E_{xc}[\rho_s(\mathbf{r},\mathbf{r}')]$, with the energy minimized with respect to the orbitals (e.g., HF and MCY3),

$$\frac{\partial E_v}{\partial N}=\varepsilon_f^{\mathrm{GKS}}, \tag{10}$$

where $\varepsilon_f^{\mathrm{GKS}}$ is the eigenvalue of the frontier orbital of the nonlocal potential $v+v_J+v_{xc}^{\mathrm{NL}}(\mathbf{r},\mathbf{r}')$ as in Eq. (6).

All three cases can be unified in the expression

$$\frac{\partial E_v}{\partial N}=\langle\phi_f|H_{\mathrm{eff}}|\phi_f\rangle, \tag{11}$$

with ϕ_f as the corresponding orbital (ϕ_f^{KS} or ϕ_f^{GKS}) and $H_{\mathrm{eff}}=-\frac{1}{2}\nabla^2+v+v_J+v_{xc}(\mathbf{r})$ for case A, where $E_{xc}=E_{xc}[\rho(\mathbf{r})]$, and $H_{\mathrm{eff}}=-\frac{1}{2}\nabla^2+v+v_J+v_{xc}^{\mathrm{NL}}(\mathbf{r},\mathbf{r}')$ for cases B and C, where $E_{xc}=E_{xc}[\rho_s(\mathbf{r}',\mathbf{r})]$. In cases A and C, $\partial E_v/\partial N$ is equal to the corresponding eigenvalue, but not in case B. The only differ-

TABLE I. Comparison of $\varepsilon_f^{\mathrm{GKS}}$ against corresponding experimental numbers for LDA, HF, and MCY3. MAE is mean absolute error and more details can be found in (Ref. 14).

	$\varepsilon_{\mathrm{LUMO}}^{\mathrm{GKS}}-\varepsilon_{\mathrm{HOMO}}^{\mathrm{GKS}}$			$I-A$	$-\varepsilon_{\mathrm{HOMO}}^{\mathrm{GKS}}$			I	$-\varepsilon_{\mathrm{LUMO}}^{\mathrm{GKS}}$			A
Mol	LDA	HF	MCY3	Expt	LDA	HF	MCY3	Expt	LDA	HF	MCY3	Expt
C	0.08	12.76	10.03	10.00	6.09	11.94	11.12	11.27	6.01	−0.82	1.09	1.27
O	0.23	16.80	11.57	12.16	7.28	14.11	13.01	13.62	7.05	−2.70	1.44	1.46
F_2	3.38	20.44	13.75	14.40	9.53	18.13	15.17	15.70	6.15	−2.31	1.42	1.30
OH	0.10	16.56	11.23	11.40	7.21	13.90	12.70	13.20	7.11	−2.67	1.47	1.80
MAE	11.043	4.650	0.360		5.920	1.073	0.448		5.123	3.583	0.163	
				Errors from the explicit calculation of I and A (Ref. 14)								
MAE	1.360	1.433	0.738		1.065	1.660	0.383		0.380	1.758	0.355	

ence between cases B and C is the orbitals used to evaluate Eq. (11).

Combining Eqs. (2) and (11) thus leads to the general expression for the band gap from derivatives for an N_0-electron system,

$$E_{\mathrm{gap}}^{\mathrm{deriv}}=\langle\phi_{N_0+1}|H_{\mathrm{eff}}|\phi_{N_0+1}\rangle-\langle\phi_{N_0}|H_{\mathrm{eff}}|\phi_{N_0}\rangle. \tag{12}$$

We consider a few illustrative atomic and molecular systems for which we have performed self-consistent calculations using a cc-pVQZ basis set in a modified version of CADPAC. We compare $-\varepsilon_{\mathrm{HOMO}}^{\mathrm{GKS}}$ with the experimental I, $-\varepsilon_{\mathrm{LUMO}}^{\mathrm{GKS}}$ with the experimental A, and also their corresponding differences. The results for LDA, HF, and MCY3 are given in Table I. MCY3 gives very good agreement between $-\varepsilon_{\mathrm{HOMO}}^{\mathrm{GKS}}$ and I, which is to be expected from its straight line behavior. We should emphasize that this has not been seen before for calculations with approximate exchange-correlation functionals. The error is relatively small, 0.5 eV, and is roughly similar in I, A, and also the difference $I-A$. LDA eigenvalues have a large error, with a consistent underestimation of I by about 5 eV, overestimation of A by about 4 eV, and a poor prediction of the gap. LDA does well for the explicit calculation of the $N+1$ and $N-1$ systems; it is just the use of the derivatives at N that leads to large errors. This is now clearly understood from the fractional charge picture. The HF $-\varepsilon_{\mathrm{HOMO}}^{\mathrm{GKS}}$ is often close to I, however, there are larger errors for A and also for the gap.

The band-gap issue is well understood for calculations with local density functionals (case A) or orbital functionals (case C). We now consider case B, orbital functionals in an OEP calculation, using the Yang-Wu direct minimization method.[19,20] In Fig. 1, the OEP minimized energy is remarkably similar to the GKS minimized energy in both integer and fractional charge systems. We would therefore expect $\partial E_v/\partial N_{N_0\pm\delta N}$ to be the same as the GKS derivatives.

Table II shows the eigenvalues from an OEP calculation using a Fermi-Amaldi base potential, which has the correct asymptotic behavior. The asymptotic form of the potential has a large effect on the OEP eigenvalues, but not on the eigenvalue differences or the energy derivatives. The OEP-(MCY3) $-\varepsilon_{\mathrm{HOMO}}^{\mathrm{KS}}\simeq I$ (as is proven for the exact functional[21]); however, the $E_{\mathrm{gap}}^{\mathrm{KS}}$ is much smaller than the exact gap. The inclusion of Δ_{xc}^f gives a much better agreement between the derivatives and the GKS eigenvalues. This brings us on to the nature of this term; it is only the difference between KS and GKS calculations and is needed to correctly give the derivative at N_0. It does not, however, address the question of whether the functional used for the calculation has the correct straight line behavior for fractional numbers of electrons, which is the key question in the evaluation of the band gap.

Figure 2 shows the behavior of the eigenvalues for carbon with different numbers of electrons using MCY3. The GKS frontier eigenvalues for a fractional system are almost constant between integers due to the straight line behavior of

TABLE II. Comparison of GKS and OEP eigenvalues and OEP energy derivative, $\partial E/\partial n_f=\varepsilon_f^{\mathrm{KS}}+\Delta_{xc}^f$, for HF and MCY3 (see Ref. 14 for more examples).

Mol		MCY3 $\varepsilon^{\mathrm{KS}}$	MCY3 $\varepsilon^{\mathrm{KS}}+\Delta_{xc}$	MCY3 $\varepsilon^{\mathrm{GKS}}$	HF $\varepsilon^{\mathrm{KS}}$	HF $\varepsilon^{\mathrm{KS}}+\Delta_{xc}$	HF $\varepsilon^{\mathrm{GKS}}$	Expt
C	$I-A$	0.70	10.16	10.03	1.47	13.49	12.76	10.00
	I	10.58	11.11	11.12	11.97	11.94	11.94	11.27
	A	9.88	0.96	1.09	10.50	−1.55	−0.82	1.27
F_2	$I-A$	4.06	13.74	13.75	5.62	20.49	20.44	14.40
	I	14.67	15.16	15.17	15.94	18.11	18.13	15.70
	A	10.61	1.42	1.42	10.32	−2.37	−2.31	1.30

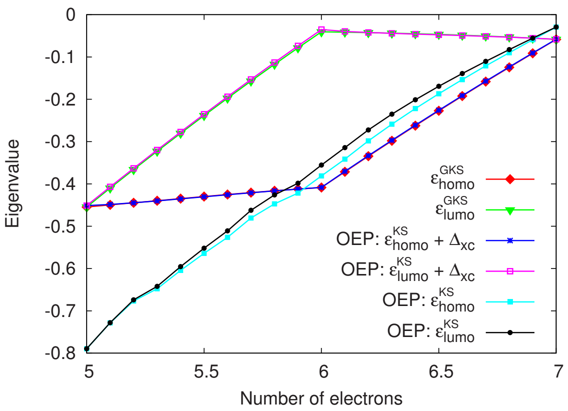

FIG. 2. (Color online) Comparison of MCY3 eigenvalues from GKS and OEP calculations and also including Δ^f_{xc} for the carbon atom.

MCY3. The OEP eigenvalues are markedly different to the GKS eigenvalues, but they become almost identical upon inclusion of Δ^f_{xc}, which can be understood from comparing Eqs. (9) and (10). The LUMO at $N_0-\delta N$ is connected to the HOMO at $N_0+\delta N$. This is clear from Eq. (7) as the change in the number of electrons N is only through the frontier occupation numbers n_f and the potential and, therefore, eigenfunctions remain fixed; i.e., there is no mysterious discontinuity in the eigenvalues.

In conclusion, we have carried out analysis and calculations on systems with fractional numbers of electrons to gain understanding of the band-gap problem in DFT. We show that the band gap is only given by the eigenvalue difference if the functional has the correct linear behavior for systems with a fractional charge. We have recently developed a functional with this linear behavior giving $-\varepsilon^{GKS}_{HOMO} \simeq I$ and $-\varepsilon^{GKS}_{LUMO} \simeq A$ and a good prediction of the band gap in molecules. We have also considered OEP calculations and showed that the derivative of the energy with respect to number of electrons is not given by the OEP eigenvalue. When the derivative is correctly evaluated, it gives practically the same as in GKS calculations. Our work thus provides the insight that it is possible to have a functional which gives the correct band gap from the eigenvalues or derivative information, so long as it has the correct fractional charge behavior and accurate energies for integer systems. Such possible functionals include explicit functionals of the electron density $E_{xc}[\rho(\mathbf{r})]$. We have only considered the explicit calculation of molecules but the same ideas are undoubtedly of key importance in solids. The understanding gained in this paper offers a different perspective and way forward for accurate calculations of the band gap in DFT.

ACKNOWLEDGMENT

Support from NSF is greatly appreciated.

APPENDIX: DETAILS OF THE DERIVATION OF $\frac{\partial E}{\partial N}$ EQUATIONS

To investigate the behavior of the total energy as a function of numbers of electrons, we consider a noninteracting ground-state reference system where we allow the occupation numbers of the orbitals to vary the number of electrons smoothly; the first-order reduced density matrix of the reference system is given by

$$\rho_s(\mathbf{r}',\mathbf{r}) = \sum_i n_i \phi_i(\mathbf{r}') \phi_i^*(\mathbf{r}), \tag{A1}$$

where $n_i=1$ for $i<f$, $n_i=\delta N$ for $i=f$, and $n_i=0$ for $i>f$, and f is the index for the frontier orbital. The electron density is just its diagonal, $\rho_s(\mathbf{r})=\rho_s(\mathbf{r},\mathbf{r})$.

In the fractional charge noninteracting system, the orbitals $\{|\phi_i\rangle\}$ are the eigenstates of a one-electron local potential $v_s(\mathbf{r})$,

$$\left(-\frac{1}{2}\nabla^2 + v_s\right)|\phi_i^{KS}\rangle = \varepsilon_i^{KS}|\phi_i^{KS}\rangle, \tag{A2}$$

or a nonlocal potential $v_s^{NL}(\mathbf{r},\mathbf{r}')$,

$$\left(-\frac{1}{2}\nabla^2 + v_s^{NL}\right)|\phi_i^{GKS}\rangle = \varepsilon_i^{GKS}|\phi_i^{GKS}\rangle. \tag{A3}$$

The former is the original KS reference system and the latter has been called the Hartree-Fock-Kohn-Sham (HFKS)[15] or the GKS method.[16]

For the Kohn-Sham reference system with local potential $v_s(\mathbf{r})$, we here use the potential functional formulation.[22] The electron density $\rho_s(\mathbf{r})$ can be represented as the set of orbitals and occupation numbers $\{\phi_i,n_i\}$ or equivalently as the local potential and total particle number $\{v_s(\mathbf{r}),N\}$. Thus, the total energy functional, formally in terms of the density as $E_v[\rho_s(\mathbf{r})]$, can be expressed as $E_v[v_s(\mathbf{r}),N]$.

The ground-state energy is the minimum of the KS energy functional, expressed (explicitly or implicitly) in terms of the local potential $v_s(\mathbf{r})$,

$$E_v(N) = \min_{v_s} E_v[v_s,N] = E_v[v_s^{gs},N], \tag{A4}$$

where the minimizer v_s^{gs} is the OEP, as established recently.[22] The variational nature of v_s^{gs} means $\frac{\delta E_v[v_s,N]}{\delta v_s(\mathbf{r})}\Big|_{v_s^{gs}}=0$, simplifying the calculation of the derivative,

$$\frac{\partial E_v}{\partial N} = \int d\mathbf{r}\, \frac{\delta E_v[v_s,N]}{\delta v_s(\mathbf{r})}\bigg|_{v_s^{gs}} \frac{\partial v_s^{gs}(\mathbf{r})}{\partial N} + \left(\frac{\partial E_v[v_s^{gs},N]}{\partial N}\right)_{v_s^{gs}}$$
$$= \left(\frac{\partial E_v[v_s^{gs},N]}{\partial N}\right)_{v_s^{gs}}. \tag{A5}$$

We now would like to express the result of Eq. (A5) in terms of $\{\phi_i,n_i\}$. Consider a change in the total number of electrons $N=N_0+\delta N$, where N_0 is an integer and $|\delta N|<1$. At fixed v_s^{gs}, all the orbitals $\{\phi_i^{v_s^{gs}}\}$, as its eigenstates, are fixed. Since $\rho_s(\mathbf{r})$ is the ground-state density of the reference potential v_s^{gs}, only the frontier level occupation n_f is allowed to change $\delta N=\delta n_f$, thus,

$$\frac{\partial E_v(N)}{\partial N} = \left(\frac{\partial E_v[\{\phi_i^{v_s^{gs}}, n_i\}]}{\partial n_f}\right)_{\{\phi_i^{v_s^{gs}}\}}, \tag{A6}$$

where the frontier orbital is either the LUMO or the HOMO,

$$\begin{aligned} n_f &= n_{\mathrm{LUMO}} \quad \text{if } \delta N > 0 \\ n_f &= n_{\mathrm{HOMO}} \quad \text{if } \delta N < 0. \end{aligned} \tag{A7}$$

Now, we apply Eq. (A6) to the KS energy,

$$\begin{aligned} E_v[\{\phi_i^{\mathrm{KS}}, n_i\}] = \sum_i -\frac{1}{2} n_i \langle \phi_i^{\mathrm{KS}} | \nabla^2 | \phi_i^{\mathrm{KS}} \rangle + J[\rho] + E_{xc}[\rho] \\ + \int v(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r}, \end{aligned} \tag{A8}$$

and obtain

$$\begin{aligned} \frac{\partial E_v}{\partial N} = -\frac{1}{2} \langle \phi_f^{\mathrm{KS}} | \nabla^2 | \phi_f^{\mathrm{KS}} \rangle + \int (\phi_f^{\mathrm{KS}}(\mathbf{r}))^* [v(\mathbf{r}) \\ + v_J(\mathbf{r})] \phi_f^{\mathrm{KS}}(\mathbf{r}) d\mathbf{r} + \left(\frac{\partial E_{xc}}{\partial n_f}\right)_{\{\phi_i^{v_s^{gs}}\}}. \end{aligned} \tag{A9}$$

That is the general result. There are two common types of E_{xc} functionals where we can obtain further details.

Case A. E_{xc} is $E_{xc}[\rho_s(\mathbf{r})]$, an explicit functional of ρ_s (e.g., LDA or GGA). Then,

$$\begin{aligned} \left(\frac{\partial E_{xc}[\rho_s(\mathbf{r})]}{\partial n_f}\right)_{\{\phi_i^{v_s^{gs}}\}} &= \int \frac{\delta E_{xc}[\rho_s(\mathbf{r})]}{\delta \rho_s(\mathbf{r})} \left(\frac{\partial \rho_s(\mathbf{r})}{\partial n_f}\right)_{\{\phi_i^{v_s^{gs}}\}} d\mathbf{r} \\ &= \int (\phi_f^{\mathrm{KS}}(\mathbf{r}))^* v_{xc}(\mathbf{r}) \phi_f^{\mathrm{KS}}(\mathbf{r}) d\mathbf{r}, \end{aligned} \tag{A10}$$

where we use the conventional local exchange-correlation potential,

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}. \tag{A11}$$

From Eq. (A9), we have

$$\begin{aligned} \frac{\partial E_v}{\partial N} = -\frac{1}{2} \langle \phi_f^{\mathrm{KS}} | \nabla^2 | \phi_f^{\mathrm{KS}} \rangle + \int (\phi_f^{\mathrm{KS}}(\mathbf{r}))^* [v(\mathbf{r}) + v_J(\mathbf{r}) \\ + v_{xc}(\mathbf{r})] \phi_f^{\mathrm{KS}}(\mathbf{r}) d\mathbf{r} = \epsilon_f^{\mathrm{KS}}, \end{aligned} \tag{A12}$$

where ϵ_f^{KS} is the KS eigenvalue for the frontier orbital in the local potential $v_s(\mathbf{r}) = v(\mathbf{r}) + v_J(\mathbf{r}) + v_{xc}(\mathbf{r})$. This is exactly the combination of Eq. (A6) with Janak's theorem for n_f.[17]

Case B. E_{xc} is $E_{xc}[\rho_s(\mathbf{r}', \mathbf{r})]$, a functional of the first-order noninteracting density matrix (e.g., OEP exact exchange). These functionals are often referred to as orbital functionals. Then,

$$\begin{aligned} \left(\frac{\partial E_{xc}[\rho_s(\mathbf{r}, \mathbf{r}')]}{\partial n_f}\right)_{\{\phi_i^{v_s^{gs}}\}} &= \int \frac{\delta E_{xc}}{\delta \rho_s(\mathbf{r}, \mathbf{r}')} \left(\frac{\partial \rho_s(\mathbf{r}, \mathbf{r}')}{\partial n_f}\right)_{\{\phi_i^{v_s^{gs}}\}} d\mathbf{r} d\mathbf{r}' \\ &= \int (\phi_f^{\mathrm{KS}}(\mathbf{r}))^* v_{xc}^{\mathrm{NL}}(\mathbf{r}, \mathbf{r}') \phi_f^{\mathrm{KS}}(\mathbf{r}') d\mathbf{r} d\mathbf{r}', \end{aligned} \tag{A13}$$

where we define the nonlocal exchange-correlation potential as

$$v_{xc}^{\mathrm{NL}}(\mathbf{r}, \mathbf{r}') = \frac{\delta E_{xc}}{\delta \rho_s(\mathbf{r}, \mathbf{r}')}. \tag{A14}$$

From Eq. (A9), we have

$$\begin{aligned} \frac{\partial E_v}{\partial N} &= -\frac{1}{2} \langle \phi_f^{\mathrm{KS}} | \nabla^2 | \phi_f^{\mathrm{KS}} \rangle \\ &\quad + \int (\phi_f^{\mathrm{KS}}(\mathbf{r}))^* [v(\mathbf{r}) + v_J(\mathbf{r})] \phi_f^{\mathrm{KS}}(\mathbf{r}) d\mathbf{r} \\ &\quad + \int (\phi_f^{\mathrm{KS}}(\mathbf{r}))^* v_{xc}^{\mathrm{NL}}(\mathbf{r}, \mathbf{r}') \phi_f^{\mathrm{KS}}(\mathbf{r}') d\mathbf{r} d\mathbf{r}' \\ &= \left\langle \phi_f^{\mathrm{KS}} \middle| -\frac{1}{2} \nabla^2 + v + v_J + v_{xc}^{\mathrm{NL}} \middle| \phi_f^{\mathrm{KS}} \right\rangle \end{aligned} \tag{A15a}$$

$$= \varepsilon_f^{\mathrm{KS}} + \langle \phi_f^{\mathrm{KS}} | v + v_J + v_{xc}^{\mathrm{NL}} - v_s | \phi_f^{\mathrm{KS}} \rangle, \tag{A15b}$$

which is a key result of this work, showing that for general orbital functionals, $\frac{\partial E_v}{\partial N}$ is not given by the frontier OEP eigenvalue $\varepsilon_f^{\mathrm{KS}}$, but with a correction term. This general result agrees with Ref. 1 in the case of the exact exchange functional and is related to the results of Ref. 18 from the self-energy.

We next consider the HFKS or GKS reference systems with nonlocal potential $v_s^{\mathrm{NL}}(\mathbf{r}, \mathbf{r}')$. In this case, like in Hartree-Fock theory, the ground-state energy is just the following minimum:

$$E_v(N) = \min_{\{\phi_i^{\mathrm{GKS}}\}} E_v[\{\phi_i^{\mathrm{GKS}}, n_i\}] = E_v[\{\phi_i^{gs}, n_i\}], \tag{A16}$$

where the minimizer ϕ_i^{gs} is the eigenstate of $v_s^{\mathrm{NL}}(\mathbf{r}, \mathbf{r}')$ as in Eq. (A3). The variational nature of the orbitals ϕ_i^{gs} also simplifies the derivative,

$$\begin{aligned} \frac{\partial E_v}{\partial N} &= \sum_i \int d\mathbf{r} \left. \frac{\delta E_v[\{\phi_i^{\mathrm{GKS}}, n_i\}]}{\delta \phi_i^{\mathrm{GKS}}(\mathbf{r})} \right|_{\{\phi_i^{gs}\}} \frac{\partial \phi_i^{gs}(\mathbf{r})}{\partial N} \\ &\quad + \left(\frac{\partial E_v[\{\phi_i^{gs}, n_i\}]}{\partial N}\right)_{\{\phi_i^{gs}\}} = \left(\frac{\partial E_v[\{\phi_i^{gs}, n_i\}]}{\partial N}\right)_{\{\phi_i^{gs}\}} \end{aligned} \tag{A17a}$$

$$= \left(\frac{\partial E_v[\{\phi_i^{gs}, n_i\}]}{\partial n_f}\right)_{\{\phi_i^{gs}\}}, \tag{A17b}$$

which is similar to Eq. (A6). Then, we have the third scenario.

Case C. E_{xc} is $E_{xc}[\rho_s(\mathbf{r}', \mathbf{r})]$, with the orbitals ϕ_i^{GKS} being the minimizer in Eq. (A17a) and thus eigenstates of v_s^{NL} [Eq. (A3)]. Then,

$$\frac{\partial E_v}{\partial N} = -\frac{1}{2}\langle \phi_f^{\mathrm{GKS}}|\nabla^2|\phi_f^{\mathrm{GKS}}\rangle + \int (\phi_f^{\mathrm{GKS}}(\mathbf{r}))^*[v(\mathbf{r}) + v_J(\mathbf{r})]\phi_f^{\mathrm{GKS}}(\mathbf{r})d\mathbf{r} + \int (\phi_f^{\mathrm{GKS}}(\mathbf{r}))^* v_{xc}^{\mathrm{NL}}(\mathbf{r},\mathbf{r}')\phi_f^{\mathrm{GKS}}(\mathbf{r}')d\mathbf{r} = \varepsilon_f^{\mathrm{GKS}}, \tag{A18}$$

where $\varepsilon_f^{\mathrm{GKS}}$ is the eigenvalue of the frontier orbital of the nonlocal potential $v+v_J+v_{xc}^{\mathrm{NL}}(\mathbf{r},\mathbf{r}')$, as ϕ_f^{GKS} is its eigenstate.

All three cases can be unified in the expression for the derivative of the total energy with N as

$$\frac{\partial E_v}{\partial N} = \langle \phi_f|H_{\mathrm{eff}}|\phi_f\rangle, \tag{A19}$$

with ϕ_f as the corresponding orbital (ϕ_f^{KS} or ϕ_f^{GKS}) and $H_{\mathrm{eff}}=-\frac{1}{2}\nabla^2+v+v_J+v_{xc}(\mathbf{r})$ for case A, when the exchange-correlation energy is known as an explicit functional of density as $E_{xc}=E_{xc}[\rho(\mathbf{r})]$, and $H_{\mathrm{eff}}=-\frac{1}{2}\nabla^2+v+v_J+v_{xc}(\mathbf{r},\mathbf{r}')$ for cases B and C, when the exchange-correlation energy is known as an explicit functional of the density matrix as $E_{xc}=E_{xc}[\rho_s(\mathbf{r}',\mathbf{r})]$, or as orbital functionals are often called. In cases A and C, $\frac{\partial E_v}{\partial N}$ is equal to the corresponding eigenvalue and not in case B. The only difference in the overall expression between cases B and C is the orbitals used to evaluate the expression.

---

[1] M. Städele, J. A. Majewski, P. Vogl, and A. Görling, Phys. Rev. Lett. **79**, 2089 (1997).

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

[3] M. Gruning, A. Marini, and A. Rubio, Phys. Rev. B **74**, 161103(R) (2006).

[4] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).

[5] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).

[6] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).

[7] Y. Zhang and W. Yang, J. Chem. Phys. **109**, 2604 (1998).

[8] P. Mori-Sánchez, A. J. Cohen, and W. Yang, J. Chem. Phys. **125**, 201102 (2006).

[9] A. Ruzsinszky, J. P. Perdew, G. I. Csonka, I. O. A. Vydrov, and G. E. Scuseria, J. Chem. Phys. **126**, 104102 (2006).

[10] N. Helbig, N. N. Lathiotakis, M. Albrecht, and E. K. U. Gross, Europhys. Lett. **77**, 67003 (2007).

[11] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Jr., Phys. Rev. Lett. **49**, 1691 (1982).

[12] W. Yang, Y. Zhang, and P. W. Ayers, Phys. Rev. Lett. **84**, 5172 (2000).

[13] A. J. Cohen, P. Mori-Sánchez, and W. Yang, J. Chem. Phys. **126**, 191109 (2007).

[14] See EPAPS Document No. E-PRBMDO-77-115807 for additional calculations on atoms and molecules. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.

[15] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).

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

[17] J. F. Janak, Phys. Rev. B **18**, 7165 (1978).

[18] M. E. Casida, Phys. Rev. B **59**, 4694 (1999).

[19] W. Yang and Q. Wu, Phys. Rev. Lett. **89**, 143002 (2002).

[20] Q. Wu and W. Yang, J. Theor. Comput. Chem. **2**, 627 (2003).

[21] J. P. Perdew and M. Levy, Phys. Rev. B **56**, 16021 (1997).

[22] W. Yang, P. W. Ayers, and Q. Wu, Phys. Rev. Lett. **92**, 146404 (2004).