

# Derivative discontinuity, bandgap and lowest unoccupied molecular orbital in density functional theory

Weitao Yang,<sup>1</sup> Aron J. Cohen,<sup>2</sup> and Paula Mori-Sánchez<sup>3</sup>

<sup>1</sup>*Department of Chemistry, Duke University, Durham, North Carolina 27708, USA and Department of Physics, Faculty of Science, King Abdulaziz University, PO Box 80203, Jeddah 21589, Saudi Arabia*

<sup>2</sup>*Department of Chemistry, Lensfield Rd, University of Cambridge, CB2 1EW, United Kingdom*

<sup>3</sup>*Department of Chemistry, Universidad Autónoma de Madrid, Madrid 28049, Spain*

(Received 20 January 2012; accepted 25 March 2012; published online 24 May 2012)

The conventional analysis of Perdew and Levy, and Sham and Schlüter shows that the functional derivative discontinuity of the exchange-correlation density functional plays a critical role in the correct prediction of bandgaps, or the chemical hardness. In a recent work by the present authors, explicit expressions for bandgap prediction with all common types of exchange-correlation functionals have been derived without invoking the concept of exchange-correlation energy functional derivative discontinuity at all. We here analyze the two approaches and establish their connection and difference. The present analysis further leads to several important results: (1) The lowest unoccupied molecular orbital (LUMO) in DFT has as much meaning in describing electron addition as the highest occupied molecular orbital (HOMO) in describing electron removal. (2) Every term in the total energy functional contributes to the energy gap because of the discontinuity of the derivative of the density (or density matrix) with respect to the number of electrons,  $((\partial \rho_s(\mathbf{r}', \mathbf{r}))/\partial N)_{v_s}$ , at integers. (3) Consistent with the Perdew-Levy-Sham-Schlüter conclusion that the exact Kohn-Sham energy gap differs from the fundamental bandgap by a finite correction due to the functional derivative discontinuity of the exchange-correlation energy, we show that the exchange-correlation functional cannot be an explicit and differentiable functional of the electron density, either local or nonlocal. The last result is further strengthened when we consider Mott insulators. There, the exact exchange-correlation functional needs to have an explicitly discontinuous (nondifferentiable) dependence on the density or the density matrix. (4) We obtain exact conditions on the derivatives of total energy with respect to the spin-up and spin-down number of electrons. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.3702391>]

## I. INTRODUCTION

Electron removal and addition are two fundamental electronic processes in chemistry and material science. For solids, the difference between the ionization potential and the electron affinity is the fundamental bandgap. For molecules, this same difference is the chemical hardness identified by Parr and Pearson,<sup>1</sup> neglecting a factor of one half. The bandgap plays a critical role in determining the properties of electron transport, structure and energetics of defects and interfaces, and many electromagnetic responses.<sup>2</sup> Hardness and chemical potential<sup>3</sup> and Fukui functions<sup>4</sup> are the key quantities of chemical reactivity theory.<sup>5–7</sup>

Kohn-Sham density functional theory (DFT) offers a practical and effective approach for electronic structure calculations of molecules and condensed matter.<sup>8</sup> The great success of DFT is that simple density functional approximations (DFA) perform remarkably well for a wide range of problems in chemistry and physics,<sup>9–11</sup> particularly for the prediction of the structure and thermodynamic properties of molecules and solids. In particular, hybrid<sup>9,10,12</sup> and screened hybrid functionals<sup>13</sup> have demonstrated significantly improved performance for band structure and bandgap prediction, over the local density approximation (LDA) and the generalized gradient approximation (GGA).

The range-separated approach with different treatments for the long- and short-range parts of electron-electron Coulomb interaction operator, originally formulated by Savin,<sup>14</sup> has led to much improved property prediction for charge transfer systems or optical responses by Hirao and co-workers<sup>15,16</sup> and Handy and co-workers.<sup>17</sup> It also motivated screened-Coulomb functionals with much improved description of band structure of solids by Scuseria and co-workers,<sup>13</sup> and functionals with minimal delocalization error (many-electron self-interaction error) by the present authors.<sup>18–20</sup> Related development has also been made by Baer and Neuhauser,<sup>21,22</sup> based on the generalized adiabatic connection developed by Yang.<sup>23</sup>

However, there still remain major challenges in the accurate prediction of bandgaps and related properties. There is no functional capable of predicting accurate energy gaps for all range of materials with small to large gaps, and no functional capable of predicting energy gaps cross the sizes of systems from atoms, molecules to bulk. These are important deficiencies which affect the application of DFT to many systems. The main problem of DFT for energy gaps can be traced to the delocalization error because common approximate functionals exhibit a convex behavior in violation of the linearity condition for fractional charges.<sup>20</sup> Thus, approximate

functionals tend to over-delocalize the added electron or hole, because they deviate from the correct linearity condition for fractional charges.<sup>24</sup> The delocalization error was originally termed many-electron self-interaction error<sup>25</sup> to link to and differentiate from the one-electron self-interaction error.<sup>26</sup> The term of many-electron self-interaction error has also been defined in Refs. 27 and 28, but it was used to describe both convex and concave behaviors of fractional charges. The concept of delocalization error captures the physical essence of the problem – it highlights the unphysical delocalization of electrons, or unphysically low energies for delocalized electrons. As discussed in our work,<sup>20</sup> there is also a clear difference between the convex  $E(N)$  behavior as the delocalization error from the concave  $E(N)$  behavior as the localization error. One can see the consequence of delocalization error not only in bandgap, added charge/hole distribution,<sup>29</sup> charge transfer,<sup>24,30,31</sup> and molecular ions,<sup>25,27,32</sup> but also in thermochemistry of molecules.<sup>33,34</sup> Progress has been recently made to correct non-empirically the convex behavior of LDA, in a scaled modified LDA (S-MLDA) approximation.<sup>35</sup> Because of the delocalization error, currently, reliable bandgap prediction is dependent on the use of many-body Green function theory, such as the GW approximation.<sup>2,36</sup>

In this paper, we focus on the issues of bandgap calculation and derivative discontinuity within DFT for a given functional. The conventional functional derivative perspective of Perdew and Levy,<sup>37,38</sup> and Sham and Schlüter<sup>39–41</sup> shows that the functional derivative discontinuity of the exchange-correlation density functional plays a critical role in the correct prediction of bandgaps, or the chemical hardness. In our recent fractional charge perspective, explicit expressions for bandgap prediction with all common types of exchange-correlation functionals have been derived, without invoking the concept of exchange-correlation energy functional derivative discontinuity at all.<sup>19</sup> This may appear confusing. We will analyze the two perspectives (Secs. III and IV A) and establish their connection and difference and offer some insight (the rest of Sec. IV).

Related to gap prediction is the issue of frontier orbitals within DFT. It is well known that the highest occupied molecular orbital (HOMO) describes the electron removal process by having its eigenvalue equal to the ionization energy when the Kohn-Sham (KS) effective local potential vanishes at infinity.<sup>5,24,42</sup> However, the lowest unoccupied molecular orbital (LUMO) has not been established to having any related meaning, although the meaning of the KS eigenvalues is of current interests.<sup>43</sup> The analysis here will establish an interpretation for the LUMO, as rigorous as that of the HOMO. Our approach, based on the fractional charge and fractional spin perspectives<sup>19,44,45</sup> is general and applicable not only to DFT but also any method whose basic variable is not the wavefunction, such as reduced density matrix theory<sup>46</sup> or theories of the many-body Green function.<sup>47</sup>

## II. THEORETICAL BACKGROUND

We start from the KS total energy as a functional of the electron density (extended to fractional number of electrons)

and the external potential  $v(\mathbf{r})$ ,

$$E_v[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] + \int v(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}, \quad (1)$$

where  $T_s$  is the noninteracting reference KS kinetic energy.<sup>8</sup> The ground state energy,  $E_v(N)$ , is the minimum

$$E_v(N) = \min_{\int \rho(\mathbf{r})d\mathbf{r}=N} E_v[\rho], \quad (2)$$

with the constraint

$$\int \rho(\mathbf{r})d\mathbf{r} = N. \quad (3)$$

The total electron number  $N$  is allowed to be fractional<sup>24</sup> as well as integer. As a basis for further development, consider the following background information:

1. The behavior of the energy  $E_v(N)$  is continuous and piecewise linear as a function of  $N$  in the grand canonical ensemble.<sup>24</sup> For a system with fractional charge  $N = N_0 + \delta$ , where  $N_0$  is an integer and  $0 \leq \delta \leq 1$ ,

$$E_v(N_0 + \delta) = (1 - \delta)E_v(N_0) + \delta E_v(N_0 + 1), \quad (4)$$

and the electron density is a linear mixture of the  $N_0$  and  $N_0 + 1$  densities

$$\rho_{N_0+\delta}(\mathbf{r}) = (1 - \delta)\rho_{N_0}(\mathbf{r}) + \delta\rho_{N_0+1}(\mathbf{r}). \quad (5)$$

The extension to fractional spins and the further combination of fractional charges and spins have been made recently,<sup>44</sup> extending the pure state proof of the linearity condition of Eq. (4).<sup>48</sup> An earlier discussion on pure states proof can be found in Ref. 30 and additional perspectives have also been given.<sup>49,50</sup> We will not use the subscript in  $\rho$  whenever this implies no confusion on the particle number.

2. The chemical potential is the derivative of the total energy with respect to particle number when the external potential is fixed,

$$\mu = \left( \frac{\partial E_v(N)}{\partial N} \right)_v. \quad (6)$$

As a consequence of the linearity condition, Eq. (4),  $\mu$  is a constant between the integers and has a derivative discontinuity at the integers

$$\begin{aligned} \mu(N) &= \begin{cases} -I(N_0) = E(N_0) - E(N_0 - 1), & \text{if } N_0 - 1 < N < N_0 \\ -A(N_0) = E(N_0 + 1) - E(N_0), & \text{if } N_0 < N < N_0 + 1 \end{cases} \end{aligned} \quad (7)$$

where  $I(N_0)$  is the ionization potential of the  $N_0$ -electron system and  $A(N_0)$  is its electron affinity.

3. The energy gap, or the fundamental bandgap, is defined as

$$\begin{aligned} E_g^{\text{integer}} &= [E_v(N_0 - 1) - E_v(N_0)] \\ &\quad - [E_v(N_0) - E_v(N_0 + 1)] = I(N_0) - A(N_0), \end{aligned} \quad (8)$$

which involves calculations at the  $N_0 - 1$ ,  $N_0$ , and  $N_0 + 1$  integer electron numbers. This gap can be also obtained from the derivative discontinuity at  $N_0$ , the difference of the derivatives to the left and the right,

$$E_g^{\text{deriv}} = \lim_{\delta \rightarrow 0} \left\{ \left. \frac{\partial E_v}{\partial N} \right|_{N_0+\delta} - \left. \frac{\partial E_v}{\partial N} \right|_{N_0-\delta} \right\}. \quad (9)$$

For the exact functional,  $E_g^{\text{integer}} = E_g^{\text{deriv}}$  due to the linear behavior of  $E_v(N)$  (Eq. (4)). For a DFA,  $E_g^{\text{deriv}}$  is a good approximation to  $E_g^{\text{integer}}$  if the DFA has a good linearity behavior with respect to electron number.<sup>19,20</sup> For finite systems, both  $E_g^{\text{integer}}$  and  $E_g^{\text{deriv}}$  can be calculated, but for bulk systems, only  $E_g^{\text{deriv}}$  is easily accessible computationally in band-structure calculations. Expressions to obtain  $E_g^{\text{deriv}}$  for different forms of functionals have been derived recently.<sup>19</sup>

4. A formal approach for the minimum principle (Eq. (2)) is to solve the Euler-Lagrange equation

$$\mu = \frac{\delta E_v[\rho]}{\delta \rho(\mathbf{r})} = \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} + \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} + v_J(\mathbf{r}) + v(\mathbf{r}), \quad (10)$$

where  $\frac{\delta J[\rho]}{\delta \rho(\mathbf{r})} = v_J(\mathbf{r})$  is the electrostatic potential,  $v$  is the external potential, and  $\mu$  is the Lagrange multiplier for the constraint in Eq. (3). The equivalence of the Lagrange multiplier  $\mu$  in Eq. (10) with the chemical potential of Eq. (6) has been established by Parr *et al.*,<sup>3</sup> also see Refs. 5 and 51. However, this approach based on the direct minimization of energy with respect to electron density is formal and not practical, because we do not yet know the explicit functional dependence of  $T_s$  on  $\rho$  and thus we do not know how to evaluate  $\frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})}$ .

5. The practical approach for obtaining the minimum is given by the Kohn-Sham equation,<sup>8</sup> where the noninteracting reference Kohn-Sham kinetic energy, only known exactly as an implicit functional of the density, is expressed as an explicit functional of the Kohn-Sham orbitals  $\{\phi_i(\mathbf{r})\}$

$$T_s[\rho] = \sum_{i\sigma} n_{i\sigma} \int \phi_{i\sigma}^*(\mathbf{r}) \left( -\frac{1}{2} \nabla^2 \right) \phi_{i\sigma}(\mathbf{r}) d\mathbf{r}, \quad (11)$$

and the noninteracting Kohn-Sham first-order spin reduced density matrix of the reference system is given by

$$\rho_s^\sigma(\mathbf{r}', \mathbf{r}) = \sum_i n_i \phi_{i\sigma}(\mathbf{r}') \phi_{i\sigma}^*(\mathbf{r}), \quad (12)$$

where  $n_i = 1$  for  $i < f$ ,  $n_i = \delta$  for  $i = f$ , and  $n_i = 0$ , for  $i > f$ , and  $f$  is the index for the frontier orbital. The physical density  $\rho^\sigma(\mathbf{r})$  is equal to the Kohn-Sham density  $\rho_s^\sigma(\mathbf{r})$ , which is the diagonal element of  $\rho_s^\sigma(\mathbf{r}', \mathbf{r})$

$$\begin{aligned} \rho^\sigma(\mathbf{r}) &= \rho_s^\sigma(\mathbf{r}) \\ &= \rho_s^\sigma(\mathbf{r}, \mathbf{r}) = \sum_i n_{i\sigma} \phi_{i\sigma}(\mathbf{r}) \phi_{i\sigma}^*(\mathbf{r}). \end{aligned} \quad (13)$$

In a general fractional charge noninteracting system, the orbitals  $\{\phi_i\}$  are the eigenstates of an one-electron lo-

cal potential  $v_s(\mathbf{r})$

$$\left( -\frac{1}{2} \nabla^2 + v_s^\sigma(\mathbf{r}) \right) |\phi_{i\sigma}\rangle = \varepsilon_{i\sigma} |\phi_{i\sigma}\rangle, \quad (14)$$

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

$$\left( -\frac{1}{2} \nabla^2 + v_s^{\sigma,\text{NL}}(\mathbf{r}, \mathbf{r}') \right) |\phi_{i\sigma}\rangle = \varepsilon_{i\sigma}^{\text{GKS}} |\phi_{i\sigma}\rangle. \quad (15)$$

The former is the original KS method, with corresponding eigenvalues  $\{\varepsilon_{i\sigma}\}$  (or equivalently  $\{\varepsilon_{i\sigma}^{\text{KS}}\}$ ), and the latter has been called the Hartree-Fock-Kohn-Sham (HFKS) (Ref. 5) or the generalized Kohn-Sham (GKS) method,<sup>52</sup> with corresponding eigenvalues  $\{\varepsilon_{i\sigma}^{\text{GKS}}\}$ .

Note that the KS kinetic energy functional  $T_s[\rho]$ , and the KS (or the GKS) equations have been extended to fractional charge and spin systems, based on the extension of the Kohn-Sham assumption that the grand-canonical-ensemble physical density can be represented by a grand-canonical-ensemble noninteracting reference system, with the same reference potential (see supplemental material in Ref. 44). Such a grand-canonical-ensemble noninteracting reference system has the kinetic energy of Eq. (11) and density of Eq. (13).

To simplify the notations, we have suppressed the spin index throughout most of our presentation. It must be understood that we imply that the orbital index  $i$  in  $\phi_i$  does contain the spin index and that the potential ( $v_s(\mathbf{r})$  or  $v_s^{\text{NL}}(\mathbf{r}, \mathbf{r}')$ ) is different for different spins. We have here explicitly indicated the spin dependence in Eqs. (11)–(15). In our previous calculations,<sup>19,44</sup> we have always used spin-density calculations, unless stated otherwise. Within spin DFT, there are two electron numbers ( $N_\alpha, N_\beta$ ) and two spin chemical potentials ( $\mu_\alpha, \mu_\beta$ ). The relation to the total number of electron is  $N = N_\alpha + N_\beta$ , and the ground state total energy is the minimum with respect to all possible spin electrons numbers

$$E_v(N) = \min_{N_\alpha + N_\beta = N} E_v(N_\alpha, N_\beta). \quad (16)$$

Correspondingly, the chemical potentials,  $\mu_\alpha = (\frac{\partial E_v(N_\alpha, N_\beta)}{\partial N_\alpha})_v$  and  $\mu_\beta = (\frac{\partial E_v(N_\alpha, N_\beta)}{\partial N_\beta})_v$  are related to the ground state chemical potential  $\mu$  of Eq. (7) as

$$\mu(N) = \begin{cases} \max(\mu_\alpha, \mu_\beta), & \text{if } N_0 - 1 < N < N_0 \\ \min(\mu_\alpha, \mu_\beta), & \text{if } N_0 < N < N_0 + 1 \end{cases}. \quad (17)$$

In this way, the ground state chemical potential is the always related to the lowest energy state of either adding an electron, or removing an electron, among the possible spin choices of the added/removed electron. The linearity condition of  $E_v(N)$  force the exact chemical potential  $\mu(N)$  always to be a constant between integers, as in Eq. (7). Only one of the spin chemical potentials ( $\mu_\alpha, \mu_\beta$ ) satisfies such a chemical potential constancy condition, as seen from Eq. (17).

To summarize, the derivative of  $E_v(N)$  with respect to electron number is the chemical potential  $\mu$ , and its derivative discontinuity at an integer  $N_0$  that gives the energy gap. There are two paths to carry out the analysis and obtain the chemical potential  $\mu$  and the gap (or the chemical hardness): The

Perdew-Levy<sup>37</sup> and Sham-Schlüter<sup>39</sup> analysis uses the Euler-Lagrange equation (10), while the approach of the present authors evaluates the derivative  $(\frac{\partial E_v}{\partial N})_v$  directly from the quantities of the non-interactive reference KS or GKS system.<sup>53</sup> These are based on two different perspectives using two sets of independent variables: (1) the electron density  $\rho$ , and (2) the electron number  $N$  and the reference potential  $v_s$  or  $v_s^{\text{NL}}$ . The former is the conventional DFT view, while the latter is the potential functional theory (PFT) formulation of DFT.<sup>54</sup>

Derivative discontinuities play a key role in understanding the gap or chemical hardness. There are two different types of derivative discontinuities involved: the functional derivative discontinuity as in  $\frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})}$ , and the ordinary partial derivative discontinuity as in  $(\frac{\partial E_v}{\partial N})_v$ . In Perdew-Levy<sup>37</sup> and Sham-Schlüter<sup>39</sup> analysis, the gap (e.g., the derivative discontinuity in  $\mu(N)$ ) is a consequence of the functional derivative discontinuity of  $\frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})}$  and  $\frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}$  because of the use of the Euler-Lagrange equation (10). On the other hand, our analysis deals directly with the partial derivative  $(\frac{\partial E_v}{\partial N})_v$  and its discontinuity (Eq. (9)), without invoking the concept of exchange-correlation energy functional derivative discontinuity at all.<sup>53</sup> These two perspectives of the gap are analyzed and compared in Secs. III–IV.

### III. PERDEW-LEVY-SHAM-SCHLÜTER ANALYSIS OF THE GAP IN TERMS OF DENSITY FUNCTIONAL DERIVATIVES

We review here the approach of the Euler-Lagrange (density functional) analysis of Perdew-Levy<sup>37</sup> and Sham-Schlüter.<sup>39</sup> The discussion in this section will be focused on spin-compensated systems, as in the original papers. Based on Eqs. (6) and (10), the functional derivative  $\frac{\delta E_v[\rho]}{\delta \rho(\mathbf{r})}$  evaluated at the solution (the ground state density), has a derivative discontinuity of the chemical potential  $\mu$ , as expressed in Eq. (7),

$$E_g^{\text{deriv}} = \lim_{\delta \rightarrow 0} \{ \mu(N_0 + \delta) - \mu(N_0 - \delta) \} \\ = \lim_{\delta \rightarrow 0} \left\{ \left. \frac{\delta E_v[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0+\delta} - \left. \frac{\delta E_v[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0-\delta} \right\}. \quad (18)$$

Given that the functional derivatives  $\frac{\delta J[\rho]}{\delta \rho(\mathbf{r})} = v_J(\mathbf{r})$  and  $\frac{\delta \int v(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}}{\delta \rho(\mathbf{r})} = v(\mathbf{r})$  are continuous at integer number of electrons, only  $T_s[\rho]$  and  $E_{xc}[\rho]$  contribute to the discontinuity of the functional derivative,

$$E_g^{\text{deriv}} = \lim_{\delta \rightarrow 0} \left\{ \left( \left. \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0+\delta} - \left. \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0-\delta} \right) + \left( \left. \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0+\delta} - \left. \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0-\delta} \right) \right\}. \quad (19)$$

$$= \lim_{\delta \rightarrow 0} \left( \left. \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0+\delta} - \left. \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0-\delta} \right) + (v_{xc}^{(+)}(\mathbf{r}) - v_{xc}^{(-)}(\mathbf{r})), \quad (20)$$

where we define the functional derivatives at each side of an integer as

$$v_{xc}^{(\pm)}(\mathbf{r}) = \lim_{\delta \rightarrow 0} \left. \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0 \pm \delta}. \quad (21)$$

This form, while correct, does not easily lead to expressions that can be calculated, just as for the parent Euler-Lagrange equation (10). It has been proven, however, for  $N_0 = 1$ ,<sup>55</sup> that the KS eigen-energy gap comes from the discontinuity in the kinetic energy term only

$$\varepsilon_{N_0+1}(N_0) - \varepsilon_{N_0}(N_0) = \lim_{\delta \rightarrow 0} \left( \left. \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0+\delta} - \left. \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0-\delta} \right). \quad (22)$$

This equality has also been used for general  $N_0$ ,<sup>37,39</sup> and the validity of the proof has been discussed by Sagvolden and Perdew.<sup>55</sup> If we assume Eq. (22), then the following expression for the gap is obtained

$$E_g^{\text{deriv}} = \varepsilon_{N_0+1}(N_0) - \varepsilon_{N_0}(N_0) + v_{xc}^{(+)}(\mathbf{r}) - v_{xc}^{(-)}(\mathbf{r}) \quad (23)$$

$$= \varepsilon_{N_0+1}(N_0) - \varepsilon_{N_0}(N_0) + \Delta_{xc}, \quad (24)$$

where the functional derivative discontinuity of the exchange-correlation potential is

$$\Delta_{xc} = v_{xc}^{(+)}(\mathbf{r}) - v_{xc}^{(-)}(\mathbf{r}). \quad (25)$$

This is the celebrated discovery of the exchange-correlation potential derivative discontinuity and its contribution to the energy gap.<sup>37,39,56</sup> Note that Eqs. (19), (22), and (23) are all defined at every point in the three-dimensional space. Note that this formulation does not allow the direct evaluation of  $\left. \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0 \pm \delta}$  if the functional is given not in terms of the density but implicitly in terms of the KS orbitals, just like the case of  $\left. \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} \right|$ . This has been evaluated in a many-body theory framework.<sup>39</sup> Calculating optimized effective potential (OEP) can directly lead to potentials,<sup>57,58</sup> which can be related to the functional derivatives  $\frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}$  and  $\frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})}$ , but up to a constant.

### IV. BANDGAP BASED ON DIRECT EVALUATION OF THE CHEMICAL POTENTIAL $(\frac{\partial E_v}{\partial N})_v$ FROM POTENTIAL FUNCTIONAL THEORY

In potential functional theory,<sup>54,59</sup> the total energy  $E_v[v_s, N]$  is a functional of the one-electron potential  $v_s$ , either local or nonlocal, of the noninteracting reference system and the total number of electrons. Making use of the clear dependence on  $N$  in PFT, the present authors recently developed a method for directly calculating the chemical potential  $(\frac{\partial E_v}{\partial N})_v$  and hence the fundamental gap.<sup>19</sup> While expressions for chemical potentials in Sec. IV A are valid for spin density functional calculations and have been used in such calculations for open-shell systems,<sup>19</sup> to make the direct comparison with the bandgap analysis of Sec. III, we have dropped the dependence on the spin in the potentials, orbitals and eigenvalues in the rest of the discussion in the paper. Thus the comparison in Sec. IV B between the two analyses is focused on



spin-compensated systems. In PFT, the total energy and hence the energy gap are invariant with respect to an arbitrary constant  $c$  in the potential  $v_s$ ; namely for any  $N$ , integer or fraction,

$$E_v[v_s, N] = E_v[v_s + c, N]. \quad (26)$$

This presents a contrast from the conventional Perdew-Levy-Sham-Schlüter analysis based on density functional derivatives, in which the constants in the exchange-correlation potential play an essential role. We will here present and extend our PFT approach, making connection and comparison with the Perdew-Levy-Sham-Schlüter analysis.

### A. Review of the expressions for the chemical potential $(\frac{\partial E_v}{\partial N})_v$

For the Kohn-Sham reference system with local potential  $v_s(\mathbf{r})$ , we use here the potential functional formulation.<sup>54,59</sup> The electron density  $\rho_s(\mathbf{r})$  can be represented as the set of orbitals and occupation numbers  $\{\phi_i, n_i\}$ , or equivalently by 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 equivalently expressed as  $E_v[v_s(\mathbf{r}), N]$ .

The following analysis reviews and synthesizes the results of the initial derivation of the chemical potential formula in Ref. 19 with further extensions.<sup>53,60</sup> The ground state energy is the minimum of the KS energy functional, expressed (explicitly or implicitly) in terms of the local potential  $v_s(\mathbf{r})$ ,

$$E_v(N) = \min_{v_s} E_v[v_s, N] = E_v[v_s^{\text{gs}}, N], \quad (27)$$

where the minimizer  $v_s^{\text{gs}}$  is the OEP, as described in Ref. 54. The variational nature of  $v_s^{\text{gs}}$  means that  $\frac{\delta E_v[v_s, N]}{\delta v_s(\mathbf{r})}|_{v_s^{\text{gs}}} = 0$ , simplifying the calculation of the derivative,

$$\begin{aligned} \left(\frac{\partial E_v}{\partial N}\right)_v &= \int d\mathbf{r} \frac{\delta E_v[v_s, N]}{\delta v_s(\mathbf{r})} \bigg|_{v_s^{\text{gs}}} \frac{\partial v_s^{\text{gs}}(\mathbf{r})}{\partial N} + \left(\frac{\partial E_v[v_s^{\text{gs}}, N]}{\partial N}\right)_{v_s^{\text{gs}}} \\ &= \left(\frac{\partial E_v[v_s^{\text{gs}}, N]}{\partial N}\right)_{v_s^{\text{gs}}}. \end{aligned} \quad (28)$$

This is a key result linking the chemical potential to quantities in the KS reference system.<sup>19</sup> We will drop the superscript  $gs$  for the ground state quantities when there is no confusion.

In the case of the GKS reference system depending on a generalized (nonlocal) OEP  $v_s^{\text{NL}}(\mathbf{r}, \mathbf{r}')$ , a derivation for chemical potential expression completely analogous to Eq. (28) has been made with the only difference of using  $v_s^{\text{NL}}(\mathbf{r}, \mathbf{r}')$  instead of the local OEP  $v_s(\mathbf{r})$ .<sup>53</sup>

We now would like to express the result of Eq. (28) in terms of  $\{\phi_i, n_i\}$  in either KS or GKS calculations. Consider a change in the total number of electrons  $N = N_0 + \delta$ . At the fixed  $v_s^{\text{gs}}/v_s^{\text{NL}}$  all the KS/GKS orbitals  $\{\phi_i^{\text{gs}}\}$  as its eigenstates are fixed. Only the frontier level occupation  $n_f$  is allowed to change  $\delta = \delta n_f$ , thus

$$\left(\frac{\partial E_v}{\partial N}\right)_v = \left(\frac{\partial E_v[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_f}\right)_{\{\phi_i^{\text{gs}}\}}, \quad (29)$$

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

$$n_f = \begin{cases} n_{\text{lumo}} & \text{if } \delta > 0 \\ n_{\text{homo}} & \text{if } \delta < 0 \end{cases}. \quad (30)$$

Now we apply Eq. (29) to the total energy

$$\begin{aligned} E_v[\{\phi_i, n_i\}] &= \sum_i -\frac{1}{2} n_i \langle \phi_i | \nabla^2 | \phi_i \rangle + J[\rho] + E_{xc}[\rho] \\ &\quad + \int v(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} \end{aligned} \quad (31)$$

and obtain

$$\begin{aligned} \left(\frac{\partial E_v}{\partial N}\right)_v &= -\frac{1}{2} \langle \phi_f | \nabla^2 | \phi_f \rangle + \int \phi_f^*(\mathbf{r}) \\ &\quad \times [v(\mathbf{r}) + v_J(\mathbf{r})] \phi_f(\mathbf{r}) d\mathbf{r} + \left(\frac{\partial E_{xc}}{\partial n_f}\right)_{\{\phi_i^{\text{gs}}\}}. \end{aligned} \quad (32)$$

This is our general result, where the exchange-correlation term may include, in general, an implicit or explicit dependence on the total number of electrons. This allows the calculation of the chemical potential for any functional, expressed as  $E_{xc}[\{\phi_i, n_i\}]$ . There are several specific types of  $E_{xc}$  functionals for which we can obtain further details.

**Case A:**  $E_{xc}$  is  $E_{xc}[\rho_s(\mathbf{r})]$ , an explicit and differentiable functional of  $\rho_s(\mathbf{r})$  (e.g., LDA, GGA or other nonlocal functionals). Then

$$\begin{aligned} \left(\frac{\partial E_{xc}[\rho_s(\mathbf{r})]}{\partial n_f}\right)_{\{\phi_i^{\text{gs}}\}} &= \int \frac{\delta E_{xc}[\rho_s(\mathbf{r})]}{\delta \rho_s(\mathbf{r})} \left(\frac{\partial \rho_s(\mathbf{r})}{\partial n_f}\right)_{\{\phi_i^{\text{gs}}\}} d\mathbf{r} \\ &= \int \phi_f^*(\mathbf{r}) v_{xc}(\mathbf{r}) \phi_f(\mathbf{r}) d\mathbf{r}, \end{aligned} \quad (33)$$

where we use the conventional local exchange correlation potential

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}. \quad (34)$$

In Eq. (34), there is no ambiguity on the possible constant in the functional derivative,<sup>61</sup> because the density variation involves particle number change. There is no finite jump in  $\frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}$  at the integer number of electrons either, because here we consider only the case where  $E_{xc}[\rho_s(\mathbf{r})]$  is an explicit and differentiable functional of  $\rho_s(\mathbf{r})$ .

From Eq. (32) we have

$$\begin{aligned} \left(\frac{\partial E_v(N)}{\partial N}\right)_v &= -\frac{1}{2} \langle \phi_f | \nabla^2 | \phi_f \rangle + \int \phi_f^*(\mathbf{r}) \\ &\quad \times [v(\mathbf{r}) + v_J(\mathbf{r}) + v_{xc}(\mathbf{r})] \phi_f(\mathbf{r}) d\mathbf{r}, \\ &= \varepsilon_f \end{aligned} \quad (35)$$

where  $\varepsilon_f$  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 our result, Eq. (29), with Janak's theorem for  $\frac{\partial E_v}{\partial n_f}$ .<sup>62</sup>

**Case B:**  $E_{xc}$  is  $E_{xc}[\rho_s(\mathbf{r}'), \mathbf{r}]$ , an explicit and differentiable functional of the first order noninteracting density matrix and the noninteracting reference system has a *local* KS

potential (e.g., the exact exchange, EXX in OEP calculations). This case is also often referred to as orbital functionals. Then

$$\begin{aligned} \left( \frac{\partial E_{xc}[\rho_s(\mathbf{r}, \mathbf{r}')] }{\partial n_f} \right)_{\{\phi_i^{\text{gs}}\}} &= \int \frac{\delta E_{xc}}{\delta \rho_s(\mathbf{r}, \mathbf{r}')} \left( \frac{\partial \rho_s(\mathbf{r}, \mathbf{r}')}{\partial n_f} \right)_{\{\phi_i^{\text{gs}}\}} d\mathbf{r} d\mathbf{r}' \\ &= \int \phi_f^*(\mathbf{r}) v_{xc}^{\text{NL}}(\mathbf{r}, \mathbf{r}') \phi_f(\mathbf{r}') d\mathbf{r} d\mathbf{r}', \end{aligned} \quad (36)$$

where we define the nonlocal exchange correlation potential as

$$v_{xc}^{\text{NL}}(\mathbf{r}, \mathbf{r}') = \frac{\delta E_{xc}}{\delta \rho_s(\mathbf{r}', \mathbf{r})}. \quad (37)$$

Again, in Eq. (37), there is no ambiguity on the possible constant in the functional derivative, because the density matrix variation involves particle number change. There is no finite jump in  $\frac{\delta E_{xc}}{\delta \rho_s(\mathbf{r}, \mathbf{r})}$  at the integer number of electrons either, because here we consider only the case where  $E_{xc}[\rho_s(\mathbf{r}, \mathbf{r}')]$  is an explicit and differentiable functional of  $\rho_s(\mathbf{r}, \mathbf{r})$ .

The present analysis can be extended to the case of general orbital functionals  $E_{xc}[\{\phi_i, n_i\}]$ , including both occupied and unoccupied orbitals, such as functionals based on the second-order perturbation theory.<sup>63</sup> In this situation the dependence on the orbital eigenvalues is formally included, as the eigenvalues depend on  $\{\phi_i, n_i\}$ . All we need to do is to use Eq. (32) and explicitly evaluate the partial derivative  $(\frac{\partial E_{xc}[\{\phi_i, n_i\}]}{\partial n_f})_{\{\phi_i\}}$ , which takes the place of  $\langle \phi_f | v_{xc}^{\text{NL}} | \phi_f \rangle$ . This situation has not been considered previously.<sup>19</sup>

From Eq. (32), we have

$$\begin{aligned} \left( \frac{\partial E_v(N)}{\partial N} \right)_v &= -\frac{1}{2} \langle \phi_f | \nabla^2 | \phi_f \rangle + \int \phi_f^*(\mathbf{r}) [v(\mathbf{r}) + v_J(\mathbf{r})] \phi_f(\mathbf{r}) d\mathbf{r} \\ &\quad + \int \phi_f^*(\mathbf{r}) v_{xc}^{\text{NL}}(\mathbf{r}, \mathbf{r}') \phi_f(\mathbf{r}') d\mathbf{r} \\ &= \langle \phi_f | -\frac{1}{2} \nabla^2 + v + v_J + v_{xc}^{\text{NL}} | \phi_f \rangle \end{aligned} \quad (38)$$

$$= \varepsilon_f + \langle \phi_f | v + v_J + v_{xc}^{\text{NL}} - v_s | \phi_f \rangle \quad (39)$$

$$= \varepsilon_f + \langle \phi_f | v_{xc}^{\text{NL}}(\mathbf{r}, \mathbf{r}') - v_{xc}(\mathbf{r}) | \phi_f \rangle \quad (40)$$

using the fact that  $\phi_f$  is the eigenstate of the OEP  $v_s$ , as in Eq. (14). We have also defined  $v_{xc}(\mathbf{r}) = v_s - v - v_J$ . Equation (39) or (40) is an important result, showing that for general orbital functionals,  $(\frac{\partial E_v(N)}{\partial N})_v$  is not only given by the frontier OEP eigenvalue,  $\varepsilon_f$ , but with a correction term.

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

$$E_v(N) = \min_{\{\phi_i\}} E_v[\{\phi_i, n_i\}] = E_v[\{\phi_i^{\text{gs}}, n_i\}], \quad (41)$$

where the minimizer  $\phi_i^{\text{gs}}$  is the eigenstate of  $v_s^{\text{NL}}(\mathbf{r}, \mathbf{r}')$  as in Eq. (15). The variational nature of the orbitals  $\phi_i^{\text{gs}}$  also sim-

plifies the derivative

$$\begin{aligned} \left( \frac{\partial E_v(N)}{\partial N} \right)_v &= \sum_i \int d\mathbf{r} \frac{\delta E_v[\{\phi_i, n_i\}]}{\delta \phi_i(\mathbf{r})} \bigg|_{\{\phi_i^{\text{gs}}\}} \frac{\partial \phi_i^{\text{gs}}(\mathbf{r})}{\partial N} \\ &\quad + \left( \frac{\partial E_v[\{\phi_i^{\text{gs}}, n_i\}]}{\partial N} \right)_{\{\phi_i^{\text{gs}}\}} \\ &= \left( \frac{\partial E_v[\{\phi_i^{\text{gs}}, n_i\}]}{\partial N} \right)_{\{\phi_i^{\text{gs}}\}} \end{aligned} \quad (42)$$

$$= \left( \frac{\partial E_v[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_f} \right)_{\{\phi_i^{\text{gs}}\}}, \quad (43)$$

which is similar to Eq. (29) where  $v_s^{\text{gs}}$  is the minimizer. An alternative derivation based on the potential functional formulation with a nonlocal one-electron potential  $v_s^{\text{NL}}(\mathbf{r}, \mathbf{r}')$  has been developed recently.<sup>53</sup> Thus, in all the cases, whether the noninteracting reference potential is local or nonlocal, we can summarize the result as

$$\left( \frac{\partial E_v}{\partial N} \right)_v = \left( \frac{\partial E_v[v_s^{\text{gs}}, N]}{\partial N} \right)_{v_s^{\text{gs}}} \quad (44)$$

$$= \left( \frac{\partial E_v[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_f} \right)_{\{\phi_i^{\text{gs}}\}}. \quad (45)$$

Here the  $v_s^{\text{gs}}$  can be the usual local KS reference potential, or the nonlocal GKS reference potential. With this result, which is consistent with the discussion on GKS following Eq. (15) we can study the third scenario.

**Case C:**  $E_{xc}$  is  $E_{xc}[\rho_s(\mathbf{r}, \mathbf{r}')]$ , an explicit and differentiable functional of the first order noninteracting density matrix, but the noninteracting reference system has a *nonlocal* GKS potential (e.g., the exact exchange functional, EXX, in Hartree-Fock calculations). Note that in GKS calculations, the orbitals  $\{\phi_i\}$  are the minimizer in Eq. (41) and thus eigenstates of the nonlocal  $v_s^{\text{NL}}$ , Eq. (15). Then, from Eq. (45)

$$\begin{aligned} \left( \frac{\partial E_v}{\partial N} \right)_v &= -\frac{1}{2} \langle \phi_f | \nabla^2 | \phi_f \rangle + \int \phi_f^*(\mathbf{r}) [v(\mathbf{r}) + v_J(\mathbf{r})] \phi_f(\mathbf{r}) d\mathbf{r} \\ &\quad + \int \phi_f^*(\mathbf{r}) v_{xc}^{\text{NL}}(\mathbf{r}, \mathbf{r}') \phi_f(\mathbf{r}') d\mathbf{r} \\ &= \varepsilon_f^{\text{GKS}}, \end{aligned} \quad (46)$$

where  $\varepsilon_f^{\text{GKS}}$  is the eigenvalue of the frontier orbital of the nonlocal potential  $v_s^{\text{NL}}(\mathbf{r}, \mathbf{r}') = v(\mathbf{r}) + v_J(\mathbf{r}) + v_{xc}^{\text{NL}}(\mathbf{r}, \mathbf{r}')$ , as  $\phi_f$  is its eigenstate. Consider a specific case when the  $E_{xc}$  is the Fock exchange functional,  $v_{xc}^{\text{NL}}(\mathbf{r}, \mathbf{r}')$  is just the Fock exchange operator and the GKS calculation is completely equivalent to the Hartree-Fock (HF) calculation. Then we have

$$\left( \frac{\partial E_v^{\text{HF}}}{\partial N} \right)_v^{(-)} = \varepsilon_{\text{HOMO}}^{\text{HF}} \quad (47)$$

and

$$\left( \frac{\partial E_v^{\text{HF}}}{\partial N} \right)_v^{(+)} = \varepsilon_{\text{LUMO}}^{\text{HF}}. \quad (48)$$

This gives a new meaning to the frontier eigenvalues of the HF theory, as the corresponding HF chemical potentials of electron removal or addition. In contrast, the Koopmans's theorem identifies the HF frontier eigenvalues as the ionization potential and electron affinity under the frozen orbital approximation. Another example is commonly used hybrid functionals<sup>9,10,12</sup> with a linear mixture of the Fock exchange with GGA exchange-correlation functionals. Calculations are normally carried out in the GKS manner with nonlocal potentials, thus Eq. (46) applies and we have

$$\left(\frac{\partial E_v^{\text{hybrid}}}{\partial N}\right)_v^{(-)} = \varepsilon_{\text{HOMO}}^{\text{hybrid}} \quad (49)$$

and

$$\left(\frac{\partial E_v^{\text{hybrid}}}{\partial N}\right)_v^{(+)} = \varepsilon_{\text{LUMO}}^{\text{hybrid}}, \quad (50)$$

The same can also be concluded for screened hybrid functionals.<sup>13</sup>

All three cases (A, B, and C) can be unified in the expression for the derivative of the total energy with respect to  $N$  as

$$\left(\frac{\partial E_v}{\partial N}\right)_v = \langle \phi_f | H_{\text{eff}} | \phi_f \rangle, \quad (51)$$

where  $H_{\text{eff}} = -\frac{1}{2}\nabla^2 + v + v_J + v_{xc}(\mathbf{r})$  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  $H_{\text{eff}} = -\frac{1}{2}\nabla^2 + v + v_J + v_{xc}(\mathbf{r}, \mathbf{r}')$  for cases B and C when the exchange-correlation energy is known as an explicit functional of the density matrix as  $E_{xc} = E_{xc}[\rho_s(\mathbf{r}', \mathbf{r})]$ , also called orbital functionals. In cases A and C,  $(\frac{\partial E_v(N)}{\partial N})_v$  is equal to the corresponding eigenvalue, but not in case B. The only difference in the overall expression between cases B and C is the orbitals which one uses to evaluate the expression.

**Case D:**  $E_{xc}$  is a functional of the density or first-order density matrix with explicit discontinuity.<sup>60</sup> All the functionals considered previously have a smooth dependence on the density matrix and therefore an implicit dependence on the number of electrons. However, it is also possible to have functionals with an explicit dependence on the number of electrons and in this case an extra term arises in the evaluation of  $(\frac{\partial E_v(N)}{\partial N})_v$  that must be evaluated once  $E_{xc}$  is known in terms of  $N$ .

## B. Expressions for the bandgap

The expressions for the derivatives  $(\frac{\partial E_v}{\partial N})_v$  that we have obtained lead (from Eq. (9)) to the following expressions for the gap

$$E_g^{\text{deriv}} = \lim_{N \rightarrow N_0} \left\{ \left( \frac{\partial E_v[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0+1}} \right)_{\{\phi_i^{\text{gs}}\}} - \left( \frac{\partial E_v[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0}} \right)_{\{\phi_i^{\text{gs}}\}} \right\} \quad (52)$$

$$= \langle \phi_{N_0+1} | -\frac{1}{2}\nabla^2 + v + v_J | \phi_{N_0+1} \rangle - \langle \phi_{N_0} | -\frac{1}{2}\nabla^2 + v + v_J | \phi_{N_0} \rangle \\ + \lim_{N \rightarrow N_0} \left\{ \left( \frac{\partial E_{xc}[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0+1}} \right)_{\{\phi_i^{\text{gs}}\}} - \left( \frac{\partial E_{xc}[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0}} \right)_{\{\phi_i^{\text{gs}}\}} \right\}. \quad (53)$$

This expression shows that there is always a derivative discontinuity

$$\lim_{N \rightarrow N_0} \left\{ \left( \frac{\partial E_{xc}[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0+1}} \right)_{\{\phi_i^{\text{gs}}\}} - \left( \frac{\partial E_{xc}[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0}} \right)_{\{\phi_i^{\text{gs}}\}} \right\},$$

which contributes to the gap, just like all the other terms in the energy. Furthermore, this term can be evaluated once we know the functional in terms of the KS or GKS orbitals. In contrast, the discontinuity in the functional derivative  $\frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} = v_{xc}(\mathbf{r})$  in Eq. (19), when it exists, is difficult to evaluate.

For the general case that the exchange-correlation energy functional is an explicit and differentiable functional of density or density matrix, then

$$E_g^{\text{deriv}} = \langle \phi_{N_0+1} | H_{\text{eff}} | \phi_{N_0+1} \rangle - \langle \phi_{N_0} | H_{\text{eff}} | \phi_{N_0} \rangle, \quad (54)$$

where  $H_{\text{eff}} = -\frac{1}{2}\nabla^2 + v + v_J + v_{xc}(\mathbf{r})$  for case A with  $E_{xc} = E_{xc}[\rho(\mathbf{r})]$ , and  $H_{\text{eff}} = -\frac{1}{2}\nabla^2 + v + v_J + v_{xc}(\mathbf{r}, \mathbf{r}')$  for

cases B and C with  $E_{xc} = E_{xc}[\rho_s(\mathbf{r}', \mathbf{r})]$ . This leads to the following simplified equations for the gap:

**Case A:** For general explicit and differentiable functionals of the density, either local or nonlocal,  $v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}$  has no discontinuity and therefore, based on Eq. (19), the gap is just given by

$$E_g^{\text{deriv}} = \lim_{\delta \rightarrow 0} \left\{ \left. \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0+\delta} - \left. \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} \right|_{N_0-\delta} \right\}. \quad (55)$$

On the other hand, based on Eqs. (9) and (35), we have

$$E_g^{\text{deriv}} = \varepsilon_{N_0+1}(N_0) - \varepsilon_{N_0}(N_0). \quad (56)$$

Comparing Eq. (55) from the conventional analysis<sup>37,39</sup> with Eq. (56) based on our analysis,<sup>19</sup> we see that it offers an independent proof of Eq. (22) for this such types of functionals.

**Case B:** Based on Eqs. (9) and (40), the gap is obtained from the difference of the OEP eigenvalues and a correction

term that just accounts for the difference between the KS and GKS approaches,

$$E_g^{\text{deriv}} = \varepsilon_{N_0+1}(N_0) - \varepsilon_{N_0}(N_0) + \{ \langle \phi_{N_0+1} | v_{xc}^{\text{NL}} - v_{xc} | \phi_{N_0+1} \rangle - \langle \phi_{N_0} | v_{xc}^{\text{NL}} - v_{xc} | \phi_{N_0} \rangle \}. \quad (57)$$

We can interpret the last term in the previous expression as the discontinuity in the exchange correlation potential  $\Delta_{xc} = (v_{xc}^{(+)}(\mathbf{r}) - v_{xc}^{(-)}(\mathbf{r}))$ . Thus

$$\Delta_{xc} = \{ \langle \phi_{N_0+1} | v_{xc}^{\text{NL}} - v_{xc} | \phi_{N_0+1} \rangle - \langle \phi_{N_0} | v_{xc}^{\text{NL}} - v_{xc} | \phi_{N_0} \rangle \} \quad (58)$$

based on comparing Eq. (57) with Eq. (23) and assuming that Eq. (22) is valid. This allows for the explicit evaluation of  $\Delta_{xc}$  from a given functional. For the case of the HF exchange functional, expression equivalent to Eq. (58) has been given.<sup>30,64,65</sup> And extension to the random phase approximation has also been made.<sup>66</sup>

**Case C:** Based on Eqs. (9) and (46), the gap is given by the difference in the GKS eigenvalues, demonstrating the main effect of the density matrix derivative discontinuity

$$E_g^{\text{deriv}} = \varepsilon_{N_0+1}^{\text{GKS}}(N_0) - \varepsilon_{N_0}^{\text{GKS}}(N_0). \quad (59)$$

This is an important result for general orbital functionals obtained by the present authors,<sup>19</sup> which does not fit in the conventional Perdew-Levy-Sham-Schlüter approach based on the functional derivative with respect to the density.<sup>37-41</sup> Equation (59) provides a rigorous foundation for the use of the GKS gap as a prediction of the fundamental gap, which is widely used in practice for calculations with orbital functionals such as hybrids<sup>9,10,12</sup> and screened hybrids.<sup>13</sup>

**Case D:** For non-differentiable functionals of the density matrix,

$$\begin{aligned} E_g^{\text{deriv}} &= \varepsilon_{N_0+1}(N_0) - \varepsilon_{N_0}(N_0) + \{ \langle \phi_{N_0+1} | v_{xc}^{\text{NL}} - v_{xc} | \phi_{N_0+1} \rangle \\ &\quad - \langle \phi_{N_0} | v_{xc}^{\text{NL}} - v_{xc} | \phi_{N_0} \rangle \} + \mathcal{D}_{xc} \\ &= \varepsilon_{N_0+1}(N_0) - \varepsilon_{N_0}(N_0) + \Delta_{xc} + \mathcal{D}_{xc}, \end{aligned} \quad (60)$$

within KS (OEP) calculations, and

$$E_g^{\text{deriv}} = \varepsilon_{N_0+1}^{\text{GKS}}(N_0) - \varepsilon_{N_0}^{\text{GKS}}(N_0) + \mathcal{D}_{xc}, \quad (61)$$

within GKS calculations.  $\mathcal{D}_{xc}$  is the contribution to the fundamental energy gap from the explicit derivative discontinuity of  $E_{xc}$ , introduced in our recent work.<sup>60</sup>

To summarize, the difference in the GKS eigenvalues captures all the discontinuity due to a change of orbitals as the HOMO GKS orbital stops being occupied and GKS LUMO orbital starts being occupied, including the smooth part ( $\Delta_{xc}$ ) of an orbital functional, whereas  $\mathcal{D}_{xc}$  accounts for the additional explicit discontinuity of the exchange-correlation term itself and also contributes to the gap. Therefore, to fully understand the nature of the derivative discontinuity, both  $\Delta_{xc}$  and  $\mathcal{D}_{xc}$  need to be considered.

### C. Interpretation on contributions to the gap

The fundamental gap is given by the derivative discontinuity of  $E_v(N)$ . Both Eq. (19) from a DFT perspective and

Eq. (52) from a PFT perspective are correct for calculating the bandgap. Compared with Eq. (19), Eq. (52) leads to a very different interpretation of what contributes to the energy gap. In Eq. (19), only the functional derivatives of  $T_s$  and  $E_{xc}$  contribute to the derivative discontinuity of  $E_v(N)$ , that is the gap. In other words, only  $T_s[\rho]$  and  $E_{xc}[\rho]$  have density functional derivative discontinuity and that fully accounts for the gap in the Perdew-Levy-Sham-Schlüter analysis. For noninteracting electrons, Eq. (22) contains a similar message.

In contrast, in Eq. (52), every term in the total energy,  $E_v[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] + \int v(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}$ , contributes to the discontinuity of the derivative with respect to  $N$ , and hence the gap. This is because the density  $\rho$  and the density matrix  $\rho_s(\mathbf{r}', \mathbf{r})$  have a derivative discontinuity at  $N_0$ , as described by two Fukui functions approaching the integer limit from both sides.<sup>4,5</sup> In considering the contribution to the gap, as in Eq. (52), we need the *frozen-orbital*, or fixed  $v_s^{\text{gs}}$  derivatives,

$$\begin{aligned} \left. \frac{\partial \rho_s(\mathbf{r}', \mathbf{r})}{\partial N} \right|_{\{v_s^{\text{gs}}\}} &= \begin{cases} \phi_{N_0}(\mathbf{r}')\phi_{N_0}^*(\mathbf{r}) & \text{if } N_0 - 1 < N < N_0 \\ \phi_{N_0+1}(\mathbf{r}')\phi_{N_0+1}^*(\mathbf{r}) & \text{if } N_0 < N < N_0 + 1 \end{cases}. \end{aligned} \quad (62)$$

In Eq. (52) this derivative discontinuity of  $\left. \frac{\partial \rho_s(\mathbf{r}', \mathbf{r})}{\partial N} \right|_{\{v_s^{\text{gs}}\}}$  plays a key role and makes *every term in the total energy to contribute to the energy gap*. Moreover, in Case D, apart from the discontinuity of the density matrix, it is necessary to include the explicit derivative discontinuity of the exchange-correlation functional.

In comparison, for the density functional derivative, as in Eq. (19), the density  $\rho(\mathbf{r})$  itself is the variable and a smooth functional of  $\rho(\mathbf{r})$ , including  $\rho(\mathbf{r})$  itself, does not have a functional derivative discontinuity. Thus only  $T_s[\rho]$  and  $E_{xc}[\rho]$  may contribute to the functional derivative discontinuity and the gap. Note that the functional derivatives are difficult to evaluate for functionals of orbitals or implicit functionals of the density such as  $T_s[\rho]$  and the exact exchange (HF) functional.

### D. Implications of the constancy condition for fractional spins on the energy derivatives

When the ground state is degenerate, it is possible to extend the energy linearity conditions of Eqs. (4) and (5).<sup>44</sup> Particularly in the case of spin state degeneracy, the fractional spin concept has been introduced to describe systems with large static correlation, or strongly correlated systems.<sup>44</sup> We further explore here the general consequences of the constancy condition for fractional spins<sup>44</sup> on the energy derivatives and its implications for the frontier eigenvalues. We start from the exact flat plane behavior of the energy with respect to the spin occupations  $E_v(N_\alpha, N_\beta)$ ,<sup>60</sup> and consider a movement on the plane  $dN_\alpha + dN_\beta = dN \geq 0$  or  $dN_\alpha + dN_\beta = dN \leq 0$ . Note that then

$$dE_v = \left( \frac{\partial E_v}{\partial N_\alpha} \right)_{v, N_\beta} dN_\alpha + \left( \frac{\partial E_v}{\partial N_\beta} \right)_{v, N_\alpha} dN_\beta. \quad (63)$$



In matrix and vector notation

$$dE_v = \begin{pmatrix} \frac{\partial E_v}{\partial N_\alpha} & \frac{\partial E_v}{\partial N_\beta} \end{pmatrix} \begin{pmatrix} dN_\alpha \\ dN_\beta \end{pmatrix}. \quad (64)$$

Note that

$$\frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (65)$$

then

$$\begin{aligned} dE_v &= \frac{1}{2} \begin{pmatrix} \frac{\partial E_v}{\partial N_\alpha} & \frac{\partial E_v}{\partial N_\beta} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} dN_\alpha \\ dN_\beta \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} \frac{\partial E_v}{\partial N_\alpha} + \frac{\partial E_v}{\partial N_\beta} & \frac{\partial E_v}{\partial N_\alpha} - \frac{\partial E_v}{\partial N_\beta} \end{pmatrix} \begin{pmatrix} dN_\alpha + dN_\beta \\ dN_\alpha - dN_\beta \end{pmatrix} \\ &= \frac{1}{2} \left( \frac{\partial E_v}{\partial N_\alpha} + \frac{\partial E_v}{\partial N_\beta} \right) (dN_\alpha + dN_\beta) \\ &\quad + \frac{1}{2} \left( \frac{\partial E_v}{\partial N_\alpha} - \frac{\partial E_v}{\partial N_\beta} \right) (dN_\alpha - dN_\beta). \end{aligned} \quad (66)$$

The constancy condition<sup>44</sup> for fractional spins can be expressed as

$$dE_v = 0 \text{ when } dN = dN_\alpha + dN_\beta = 0. \quad (67)$$

Note that the range of the possible changes are constrained such that  $-2S \leq N_\alpha - N_\beta \leq 2S$ , where  $S$  is the total spin. Thus when  $dN = dN_\alpha + dN_\beta = 0$ ,  $dN_\beta = -dN_\alpha$

$$dE_v = \frac{1}{2} \left( \frac{\partial E_v}{\partial N_\alpha} - \frac{\partial E_v}{\partial N_\beta} \right) 2dN_\alpha \quad (68)$$

$$= \left( \frac{\partial E_v}{\partial N_\alpha} - \frac{\partial E_v}{\partial N_\beta} \right) dN_\alpha \quad (69)$$

$$= 0. \quad (70)$$

Thus, we have

$$\left( \frac{\partial E_v}{\partial N_\alpha} \right)_{v, N_\beta} = \left( \frac{\partial E_v}{\partial N_\beta} \right)_{v, N_\alpha}. \quad (71)$$

And when  $dN \neq 0$ , we have

$$dE_v = \frac{1}{2} \left( \frac{\partial E_v}{\partial N_\alpha} + \frac{\partial E_v}{\partial N_\beta} \right) dN, \quad (72)$$

$$dE_v = \left( \frac{\partial E_v}{\partial N_\alpha} \right)_{v, N_\beta} dN \quad (73)$$

$$= \left( \frac{\partial E_v}{\partial N_\beta} \right)_{v, N_\alpha} dN, \quad (74)$$

thus

$$\left( \frac{\partial E_v}{\partial N} \right)_v = \left( \frac{\partial E_v}{\partial N_\beta} \right)_{v, N_\alpha} = \left( \frac{\partial E_v}{\partial N_\alpha} \right)_{v, N_\beta}. \quad (75)$$

This is an important result for understanding the gaps of open shell systems where the energy constancy condition implies that they are related to Mott insulators.<sup>24</sup>

## E. Further connection between the two analysis

We start from our general Eq. (52) for the energy gap

$$\begin{aligned} E_g^{\text{deriv}} &= \lim_{N \rightarrow N_0} \left\{ \left( \frac{\partial E_v[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0+1}} \right)_{\{\phi_i^{\text{gs}}\}} - \left( \frac{\partial E_v[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0}} \right)_{\{\phi_i^{\text{gs}}\}} \right\} \\ &= \langle \phi_{N_0+1} | -\frac{1}{2} \nabla^2 + v + v_J | \phi_{N_0+1} \rangle - \langle \phi_{N_0} | -\frac{1}{2} \nabla^2 + v + v_J | \phi_{N_0} \rangle \\ &\quad + \lim_{N \rightarrow N_0} \left\{ \left( \frac{\partial E_{xc}[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0+1}} \right)_{\{\phi_i^{\text{gs}}\}} - \left( \frac{\partial E_{xc}[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0}} \right)_{\{\phi_i^{\text{gs}}\}} \right\}, \end{aligned} \quad (76)$$

which has lead to the detailed development presented above for various types of functionals. If we *assume* that we can take the functional derivatives of  $E_{xc}[\{\phi_i^{\text{gs}}, n_i\}]$  with respect to the density  $\rho$ , as in the analysis of Perdew-Levy and Sham-Schlüter, we can make further connections. For smooth functionals of the density or density matrix,

$$\begin{aligned} \left( \frac{\partial E_{xc}[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0+1}} \right)_{\{\phi_i^{\text{gs}}\}} &= \lim_{N \rightarrow N_0^+} \int \frac{\delta E_{xc}}{\delta \rho_s(\mathbf{r})} \left( \frac{\partial \rho_s(\mathbf{r})}{\partial n_{N_0+1}} \right)_{\{\phi_i^{\text{gs}}\}} d\mathbf{r} \\ &= \int \phi_{N_0+1}^*(\mathbf{r}) v_{xc}^{(+)}(\mathbf{r}) \phi_{N_0+1}(\mathbf{r}) d\mathbf{r}, \end{aligned} \quad (77)$$

and

$$\begin{aligned} \left( \frac{\partial E_{xc}[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{N_0}} \right)_{\{\phi_i^{\text{gs}}\}} &= \lim_{N \rightarrow N_0^-} \int \frac{\delta E_{xc}}{\delta \rho_s(\mathbf{r})} \left( \frac{\partial \rho_s(\mathbf{r})}{\partial n_{N_0}} \right)_{\{\phi_i^{\text{gs}}\}} d\mathbf{r} \\ &= \int \phi_{N_0}^*(\mathbf{r}) v_{xc}^{(-)}(\mathbf{r}) \phi_{N_0}(\mathbf{r}) d\mathbf{r}. \end{aligned} \quad (78)$$

Then

$$\begin{aligned} E_g^{\text{deriv}} &= \langle \phi_{N_0+1} | -\frac{1}{2} \nabla^2 + v + v_J + v_{xc}^{(+)}(\mathbf{r}) | \phi_{N_0+1} \rangle \\ &\quad - \langle \phi_{N_0} | -\frac{1}{2} \nabla^2 + v + v_J + v_{xc}^{(-)}(\mathbf{r}) | \phi_{N_0} \rangle, \end{aligned} \quad (79)$$

which can be rearranged to

$$\begin{aligned}
 E_g^{\text{deriv}} &= \langle \phi_{N_0+1} | -\frac{1}{2}\nabla^2 + v + v_J + v_{xc}^{(-)}(\mathbf{r}) | \phi_{N_0+1} \rangle \\
 &\quad - \langle \phi_{N_0} | -\frac{1}{2}\nabla^2 + v + v_J + v_{xc}^{(-)}(\mathbf{r}) | \phi_{N_0} \rangle \\
 &\quad + \langle \phi_{N_0+1} | v_{xc}^{(+)}(\mathbf{r}) - v_{xc}^{(-)}(\mathbf{r}) | \phi_{N_0+1} \rangle \\
 &= \varepsilon_{N_0+1}(N_0) - \varepsilon_{N_0}(N_0) + \langle \phi_{N_0+1} | v_{xc}^{(+)}(\mathbf{r}) \\
 &\quad - v_{xc}^{(-)}(\mathbf{r}) | \phi_{N_0+1} \rangle, \quad (80)
 \end{aligned}$$

where we have assumed that the KS eigenvalues and eigenorbitals are obtained from a KS Hamiltonian with an exchange-correlation potential from the electron-deficient side,  $-\frac{1}{2}\nabla^2 + v + v_J + v_{xc}^{(-)}(\mathbf{r})$ . This result is consistent with Eq. (23).

Furthermore, this allows us to make the connection that in case of functional of the KS orbitals or KS density matrix,

$$\begin{aligned}
 \langle \phi_{N_0+1} | v_{xc}^{(+)}(\mathbf{r}) - v_{xc}^{(-)}(\mathbf{r}) | \phi_{N_0+1} \rangle &= \langle \phi_{N_0+1} | v_{xc}^{\text{NL}} - v_{xc} | \phi_{N_0+1} \rangle \\
 &\quad - \langle \phi_{N_0} | v_{xc}^{\text{NL}} - v_{xc} | \phi_{N_0} \rangle. \quad (81)
 \end{aligned}$$

If in the OEP calculation,  $v_s(\mathbf{r})$  is forced to have the correct long-range behavior and the value  $v_s(\infty)$  is consistent with the orbital functional used to derive  $v_{xc}^{\text{NL}}(\mathbf{r}, \mathbf{r})$ , then  $v_{xc} = v_{xc}^{(-)}$ , and  $\langle \phi_{N_0} | v_{xc}^{\text{NL}} - v_{xc} | \phi_{N_0} \rangle = 0$ ,<sup>19</sup> and we have

$$\langle \phi_{N_0+1} | v_{xc}^{(+)}(\mathbf{r}) - v_{xc}^{(-)}(\mathbf{r}) | \phi_{N_0+1} \rangle = \langle \phi_{N_0+1} | v_{xc}^{\text{NL}} - v_{xc} | \phi_{N_0+1} \rangle, \quad (82)$$

giving an explicit formula to calculate  $\langle \phi_{N_0+1} | v_{xc}^{(+)}(\mathbf{r}) - v_{xc}^{(-)}(\mathbf{r}) | \phi_{N_0+1} \rangle$ , which generalizes the results of Refs. 30, 64, 65, and 67 for exact exchange expressions.

## F. Comparison to the experimental bandgap

We focus now on the question: Can one use gaps calculated in DFT to compare with fundamental gaps from experiments? This depends on the type of gap, KS or GKS, and the form of the approximate functionals. The discussion of the bandgap in the DFT literature has been mostly based on the Perdew-Levy-Sham-Schlüter analysis, given by

$$E_g^{\text{deriv}} = \varepsilon_{N_0+1}(N_0) - \varepsilon_{N_0}(N_0) + \Delta_{xc}. \quad (83)$$

This equation can be understood to state that even for an exact Kohn-Sham calculation giving the exact KS gap there is a missing derivative discontinuity term. In this understanding, a Kohn-Sham calculation with any approximate functional, up to and including the exact functional, does not give the experimental gap from its eigenvalues, an additional term is needed. This perspective, while correct, does not give information on how to understand and view the all important functional derivative discontinuity,  $\Delta_{xc}$ . The existence of  $\Delta_{xc}$  has also been used to argue against the comparison of KS gaps from approximate functional calculations with experimental gaps; one often hears such objections in discussion during conferences. Whether such an argument is justified or not depends on the type of functionals used to get the KS gap.

To address this, it is appropriate to formulate the following question: given an explicit form for  $E_{xc}$ , what is the prediction of the gap? This question has been answered for all general functional forms in our work as described in Sec. IV B.

For an LDA, GGA, or any explicit and differentiable functional of density (Case A), KS calculations can accurately predict  $\Delta\varepsilon$ , but yield a vanishing  $\Delta_{xc}$ . Therefore, with such functionals, the KS gap obtained is indeed the prediction of the fundamental gap. The missing  $\Delta_{xc}$  is simply zero. For functionals that go beyond smooth functionals of the density, such as explicit and differentiable orbital functionals (Case B), it is possible to understand and calculate the  $\Delta_{xc}$  coming from  $E_{xc}$  itself using Eq. (57), once the explicit functional form is known. Or one can directly carry out GKS calculations for such orbital functionals (Case C) and obtain the GKS gap as the prediction of fundamental gap, Eq. (59). Therefore, with such functionals, the fundamental gap predicted is *not* given by the KS gap through an OEP calculation, because the correction is non-zero. But the fundamental gap predicted can be given directly by the GKS gap. For functionals with explicit discontinuity in the derivative with respect to  $N$  (Case D), the additional term  $\mathcal{D}_{xc}$  is needed. Then the KS gap is *not* the predicted fundamental gap either.

Note that for most approximate functionals the fundamental gap so predicted is often of limited use to understand the experimental gap, because the straight line behavior for the energy with respect to  $N$  is not obeyed. Only approximate functionals that have the good linear  $E$  vs  $N$  behavior directly give a  $\Delta_{xc}$  that corrects the KS gap, or a GKS gap, which is accurate when compared directly with experiment, as demonstrated for finite systems, first by the present authors,<sup>19</sup> and subsequently by Baer and co-workers<sup>68,69</sup> and Hirao and coworkers.<sup>70</sup> A non-empirical scaling correction has also been developed recently to correct the convex  $E(N)$  behavior of common approximations and a resulting S-MLDA has been shown to be capable of predicting gaps of systems from atoms, molecules, nanoparticles to bulk solids.<sup>35</sup>

## V. THE RIGHT FUNCTIONAL FORM FOR $E_{xc}$

The previous understanding of the energy gap offers insight on the form of the exchange-correlation functional. There is a large amount of molecular and atomic data on  $I$  and  $A$  relating to the KS HOMO and LUMO orbital energies from approximate functionals (some can be found in Ref. 19). In general, they are not in agreement with each other, as would be required by our result of Eq. (35), if the exchange correlation functional is an explicit and differentiable functional of the density, either local or nonlocal (Case A). The comparison of experimental bandgaps with the KS gaps also show that they are significantly different.<sup>40,41</sup> Most approximate functionals exhibit an incorrect nonlinear  $E_v(N)$  as a function of  $N$  for fractional charges that leads to delocalization error in solids.<sup>20</sup> Due to this incorrect nature of the curve, the frontier eigenvalues are not meaningful approximations to  $-I$  and  $-A$ ;  $I$  being consistently underestimated,  $A$  overestimated, and the gap largely underestimated. Furthermore, if the exact KS eigenvalues are obtained from a high-quality density via the Zhao-Morrison-Parr<sup>71</sup> or the Wu-Yang<sup>58</sup> method, they still

do not agree with the experimental  $I$  and  $A$ . The spectrum may be shifted to give  $\varepsilon_{\text{HOMO}}^{\text{KS}} = -I$ , but the LUMO eigenvalue so obtained is limited in its meaning. It has been observed that the accurate KS HOMO LUMO gaps, obtained from accurate electron densities, track the excitation energies<sup>72,73</sup> for atoms, but not for molecules.<sup>74</sup> Effort has been made to reduce the problem for hardness calculation within such functionals.<sup>75</sup>

Thus, both molecular and bulk data clearly indicate that  $E_g^{\text{deriv}} = \varepsilon_{N_0+1}(N_0) - \varepsilon_{N_0}(N_0) + \Delta_{xc}$ , with nonzero  $\Delta_{xc}$ . In the Perdew-Levy-Sham-Schlüter analysis,<sup>37,39</sup> this clearly means that there is a derivative discontinuity in exact  $E_{xc}$ . Consistently, in our analysis, the assumption leading to Eq. (56) is clearly false for the exact  $E_{xc}$ . Therefore, the exact  $E_{xc}$  cannot be an explicit and differentiable functional of the density, local or nonlocal. This is consistent with Kohn's view<sup>56</sup> that  $E_{xc}$  cannot have a regular dependence on the electron density.

On the other hand, orbital functionals, e.g., explicit and differentiable functionals of the KS/GKS orbitals, or the KS/GKS density matrix, may be sufficient to deliver accurate gaps for many molecules and bulk systems if they have the correct linear behavior for  $E_v(N)$  with respect to  $N$ .<sup>19</sup> Moreover, OEP calculations (e.g., KS calculations) with these functionals (either with a linear or nonlinear behavior for fractional charges), are able to give  $\varepsilon_{\text{HOMO}}^{\text{KS}} = -I$  approximately if the KS potential  $v_s(\mathbf{r})$  has the correct  $1/r$  long-range behavior (see supplementary information of Ref. 19), but the LUMO KS (OEP) eigenvalue has a large error. Again the spectrum can be shifted but the gap is always wrong.

However, differentiable functionals of the density matrix still dramatically fail to describe the gap of strongly correlated systems with fractional spins, such as molecules at dissociation limit or Mott insulators.<sup>60</sup> Mott insulators were also used to understand and illustrate the bandgap problem within the KS context.<sup>24,30,37</sup> In this situation (Case D) there is no contribution to the gap from the GKS frontier orbital eigenvalues, and all the gap is in the explicit discontinuity  $\mathcal{D}_{xc}$ , which is missing from all approximate functionals in the literature. This analysis clearly leads to the conclusion that the exact  $E_{xc}$  functional cannot be an explicit and differentiable functional of the density matrix  $\rho_s(\mathbf{r}', \mathbf{r})$ . It should have an *explicit discontinuity* in the derivative with respect to  $N$ .

## VI. THE MEANING OF THE LOWEST UNOCCUPIED MOLECULAR ORBITALS

The meaning of the one-electron orbitals has been traditionally related to their eigenvalues, as shown by Koopmans' theorem in Hartree-Fock theory,  $\varepsilon_{\text{HOMO}}^{\text{HF}} \approx -I$ . In DFT, however, there has been much discussion on the meaning of the orbital energies.<sup>76-78</sup> The HOMO of a system at  $N_0$ , an integer number of electrons, has been related to its corresponding ionization energy  $I$  and given its precise physical meaning,  $\varepsilon_{\text{HOMO}}^{\text{KS}} = -I$  in exact KS theory when the Kohn-Sham effective local potential vanishes at infinity.<sup>5,24</sup> However, the meaning of the LUMO at  $N_0$  has not been considered in the same way, with no fundamental connection until our work.<sup>19</sup>

Based on our previous analysis, we can now show that both HOMO and LUMO have the similar meaning in DFT concerning the electron removal and electron addition process. Starting from our general consideration for the  $(\frac{\partial E_v(N)}{\partial N})_v$ , Eq. (32), we can explicitly write for the two derivatives

$$\left(\frac{\partial E_v(N)}{\partial N}\right)_v^{(-)} = -\frac{1}{2}\langle\phi_{\text{HOMO}}|\nabla^2|\phi_{\text{HOMO}}\rangle + \int \phi_{\text{HOMO}}^*(\mathbf{r})[v(\mathbf{r}) + v_J(\mathbf{r})]\phi_{\text{HOMO}}(\mathbf{r}) d\mathbf{r} + \left(\frac{\partial E_{xc}}{\partial n_{\text{HOMO}}}\right)_{\{\phi_i\}}. \quad (84)$$

and

$$\left(\frac{\partial E_v(N)}{\partial N}\right)_v^{(+)} = -\frac{1}{2}\langle\phi_{\text{LUMO}}|\nabla^2|\phi_{\text{LUMO}}\rangle + \int \phi_{\text{LUMO}}^*(\mathbf{r})[v(\mathbf{r}) + v_J(\mathbf{r})]\phi_{\text{LUMO}}(\mathbf{r}) d\mathbf{r} + \left(\frac{\partial E_{xc}}{\partial n_{\text{LUMO}}}\right)_{\{\phi_i\}}. \quad (85)$$

Here the parallel between the HOMO in electron removal and the LUMO in electron addition is clear: the HOMO and LUMO play the same role in each corresponding process, either the reference system is the KS system with local one-electron potential or the GKS system with nonlocal one-electron potential.

Consider, specifically, Eq. (29) for the scenario of explicit and differentiable density matrix functionals (Eq. (46)), we have

$$\left(\frac{\partial E_v(N)}{\partial N}\right)_v^{(-)} = \langle\phi_{\text{HOMO}}|H_{\text{eff}}|\phi_{\text{HOMO}}\rangle,$$

and

$$\left(\frac{\partial E_v(N)}{\partial N}\right)_v^{(+)} = \langle\phi_{\text{LUMO}}|H_{\text{eff}}|\phi_{\text{LUMO}}\rangle,$$

for both KS (OEP) orbitals and the GKS orbitals. Within the PFT perspective, the OEP potential is not defined up to a constant, Eq. (26), and one can force the HOMO orbital energy to be that of  $-I$  by fixing that constant. The LUMO energy is then not directly related to  $-A$ . But within GKS, we have

$$\left(\frac{\partial E_v(N)}{\partial N}\right)_v^{(-)} = \lim_{N \rightarrow N_0} \left(\frac{\partial E_v[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{\text{HOMO}}}\right)_{\{\phi_i\}} = \varepsilon_{\text{HOMO}}^{\text{GKS}}, \quad (86)$$

$$\left(\frac{\partial E_v(N)}{\partial N}\right)_v^{(+)} = \lim_{N \rightarrow N_0} \left(\frac{\partial E_v[\{\phi_i^{\text{gs}}, n_i\}]}{\partial n_{\text{LUMO}}}\right)_{\{\phi_i\}} = \varepsilon_{\text{LUMO}}^{\text{GKS}}. \quad (87)$$

The same applies (from Eq. (35)) to explicit and differentiable functionals of the density in terms of KS eigenvalues,

$$\left(\frac{\partial E_v(N)}{\partial N}\right)_v^{(-)} = \lim_{N \rightarrow N_0} \left(\frac{\partial E_v[\rho(\mathbf{r})]}{\partial n_{\text{HOMO}}}\right)_{\{\phi_i\}} = \varepsilon_{\text{HOMO}}^{\text{KS}}. \quad (88)$$

$$\left(\frac{\partial E_v(N)}{\partial N}\right)_v^{(+)} = \lim_{N \rightarrow N_0} \left(\frac{\partial E_v[\rho(\mathbf{r})]}{\partial n_{\text{LUMO}}}\right)_{\{\phi_i\}} = \varepsilon_{\text{LUMO}}^{\text{KS}}. \quad (89)$$

Thus the KS or the GKS HOMO and LUMO eigenvalues are the corresponding chemical potentials predicted by the (approximate) functionals used. If the approximate energy functionals have the correct straight linear behavior with respect to electron number (as dictated by Eq. (2)), then  $E_v(N)$  is linear from  $N_0$  in both the electron-deficient and electron-rich sides and the chemical potentials at  $N_0$  are  $-A$  and  $-I$ , respectively, as expressed in Eq. (7). However, the discussion in Sec. V shows that the exact  $E_{xc}$  cannot be an explicit and differentiable functional of the electron density, either local or nonlocal. Furthermore, for strongly correlated systems such as Mott insulators, the exact functional cannot be an explicit and differentiable functional of the KS/GKS density matrix, it should have explicit derivative discontinuity. Such explicit derivative discontinuity should be added in all the expressions, as in Eqs. (60) and (61) for the fundamental gap.

The basic expressions for electron removal and addition involve HOMO and LUMO in the same way, and clearly show that either in KS or GKS the LUMO and its eigenvalue are as meaningful for describing electron addition as those of the HOMO for describing electron removal.

## VII. SUMMARY

The behavior of the total energy with respect to number of electrons  $E_v(N)$ , its derivatives  $\frac{\partial E_v}{\partial N}$ , and the derivative discontinuity at integer number of electrons have been analyzed from two different perspectives using different basic variables: (1) Perdew-Levy and Sham-Schlüter analysis based on the electron density as the basic variable; and (2) the understanding offered by the present authors based on the electron number  $N$  and the reference KS potential  $v_s$  with the PFT perspective. The path (1) leads to the conclusion that only  $T_s$  and  $E_{xc}$  may contribute to the gap,  $E_g^{\text{deriv}} = \varepsilon_{N_0+1}(N_0) - \varepsilon_{N_0}(N_0) + v_{xc}^{(+)}(\mathbf{r}) - v_{xc}^{(-)}(\mathbf{r})$ , which relies on a derivative discontinuity of the exchange-correlation potential at all points in space. In the second perspective, the KS potential is fixed and only the frontier orbitals change, which leads to expressions for the gap in terms of the occupation of the frontier orbitals. In this case, all the terms in the energy contribute to the gap because the density matrix itself has a derivative discontinuity at integer numbers of electrons. It is shown that the exact  $E_{xc}$  cannot be an explicit and differentiable functional of the density, local or nonlocal, and the importance of exchange-correlation functionals with a discontinuity that goes beyond the level of orbital functionals is highlighted. This is especially true for Mott insulators such as open-shell atoms where the gap comes down to a change in the exchange-correlation energy when passing through the integer that is not captured by any change in the orbitals. Finally, the understanding of the derivatives presented here shows that in DFT the LUMO and its eigenvalue are just as meaningful in describing electron addition as those of the HOMO in describing electron removal.

## ACKNOWLEDGMENTS

Support from the Office of Naval Research (N00014-09-0576) and National Science Foundation (NSF) (CHE-09-11119) is greatly appreciated (W.Y.). Additional support is also acknowledged from the Royal Society (A.J.C.) and Ramón y Cajal (P.M.S.).

- <sup>1</sup>R. G. Parr and R. G. Pearson, *J. Am. Chem. Soc.* **105**, 7512 (1983).
- <sup>2</sup>G. Onida, L. Reining, and A. Rubio, *Rev. Mod. Phys.* **74**, 601 (2002).
- <sup>3</sup>R. G. Parr, R. A. Donnelly, M. Levy, and W. E. Palke, *J. Chem. Phys.* **68**, 3801 (1978).
- <sup>4</sup>R. G. Parr and W. T. Yang, *J. Am. Chem. Soc.* **106**, 4049 (1984).
- <sup>5</sup>R. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).
- <sup>6</sup>P. Geerlings, F. De Proft, and W. Langenaeker, *Chem. Rev.* **103**, 1793 (2003).
- <sup>7</sup>*Chemical Reactivity Theory: A Density Functional View*, edited by P. K. Chattaraj (CRC, 2009).
- <sup>8</sup>W. Kohn and L. Sham, *Phys. Rev.* **140**, A1133 (1965).
- <sup>9</sup>A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- <sup>10</sup>C. T. Lee, W. T. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- <sup>11</sup>J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- <sup>12</sup>C. Adamo and V. Barone, *J. Chem. Phys.* **110**, 6158 (1999).
- <sup>13</sup>J. Heyd, G. E. Scuseria, and M. Ernzerhof, *J. Chem. Phys.* **118**, 8207 (2003).
- <sup>14</sup>A. Savin, *Recent Developments and Applications of Modern Density Functional Theory* (Elsevier, Amsterdam, 1996), p. 327.
- <sup>15</sup>H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, *J. Chem. Phys.* **115**, 3540 (2001).
- <sup>16</sup>J. W. Song, M. A. Watson, H. Sekino, and K. Hirao, *J. Chem. Phys.* **129**, 024117 (2008).
- <sup>17</sup>T. Yanai, D. P. Tew, and N. C. Handy, *Chem. Phys. Lett.* **393**, 51 (2004).
- <sup>18</sup>A. J. Cohen, P. Mori-Sanchez, and W. T. Yang, *J. Chem. Phys.* **126**, 191109 (2007).
- <sup>19</sup>A. J. Cohen, P. Mori-Sanchez, and W. T. Yang, *Phys. Rev. B* **77**, 115123 (2008).
- <sup>20</sup>P. Mori-Sanchez, A. J. Cohen, and W. T. Yang, *Phys. Rev. Lett.* **100**, 146401 (2008).
- <sup>21</sup>R. Baer and D. Neuhauser, *Phys. Rev. Lett.* **94**, 043002 (2005).
- <sup>22</sup>R. Baer, E. Livshits, and D. Neuhauser, *Chem. Phys.* **329**, 266 (2006).
- <sup>23</sup>W. T. Yang, *J. Chem. Phys.* **109**, 10107 (1998).
- <sup>24</sup>J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, *Phys. Rev. Lett.* **49**, 1691 (1982).
- <sup>25</sup>P. Mori-Sanchez, A. J. Cohen, and W. T. Yang, *J. Chem. Phys.* **125**, 201102 (2006).
- <sup>26</sup>J. P. Perdew and A. Zunger, *Phys. Rev. B* **23**, 5048 (1981).
- <sup>27</sup>A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov, and G. E. Scuseria, *J. Chem. Phys.* **125**, 194112 (2006).
- <sup>28</sup>A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov, and G. E. Scuseria, *J. Chem. Phys.* **126**, 104102 (2007).
- <sup>29</sup>A. J. Cohen, P. Mori-Sanchez, and W. T. Yang, *Science* **321**, 792 (2008).
- <sup>30</sup>J. P. Perdew, *Density Functional Methods in Physics* (Plenum, 1985), p. 265.
- <sup>31</sup>A. Dreuw and M. Head-Gordon, *J. Am. Chem. Soc.* **126**, 4007 (2004).
- <sup>32</sup>Y. K. Zhang and W. T. Yang, *J. Chem. Phys.* **109**, 2604 (1998).
- <sup>33</sup>E. R. Johnson, P. Mori-Sanchez, A. J. Cohen, and W. T. Yang, *J. Chem. Phys.* **129**, 204112 (2008).
- <sup>34</sup>T. Heaton-Burgess and W. T. Yang, *J. Chem. Phys.* **132**, 234113 (2010).
- <sup>35</sup>X. Zheng, A. J. Cohen, P. Mori-Sanchez, X. Q. Hu, and W. T. Yang, *Phys. Rev. Lett.* **107**, 026403 (2011).
- <sup>36</sup>F. Bechstedt, F. Fuchs, and G. Kresse, *Phys. Status Solidi B* **246**, 1877 (2009).
- <sup>37</sup>J. P. Perdew and M. Levy, *Phys. Rev. Lett.* **51**, 1884 (1983).
- <sup>38</sup>J. P. Perdew, *Int. J. Quantum Chem.* **28**, 497–523 (1985).
- <sup>39</sup>L. J. Sham and M. Schlüter, *Phys. Rev. Lett.* **51**, 1888 (1983).
- <sup>40</sup>R. W. Godby, M. Schlüter, and L. J. Sham, *Phys. Rev. Lett.* **56**, 2415 (1986).
- <sup>41</sup>R. W. Godby, M. Schlüter, and L. J. Sham, *Phys. Rev. B* **37**, 10159 (1988).
- <sup>42</sup>M. Levy, J. P. Perdew, and V. Sahni, *Phys. Rev. A* **30**, 2745 (1984).
- <sup>43</sup>O. V. Gritsenko, B. Braida, and E. J. Baerends, *J. Chem. Phys.* **119**, 1937 (2003).



- <sup>44</sup>A. J. Cohen, P. Mori-Sanchez, and W. T. Yang, *J. Chem. Phys.* **129**, 121104 (2008).
- <sup>45</sup>A. J. Cohen, P. Mori-Sanchez, and W. T. Yang, *Chem. Rev.* **112**, 289 (2012).
- <sup>46</sup>H. V. Aggelen, P. Bultinck, B. Verstichel, D. V. Neck, and P. W. Ayers, *Phys. Chem. Chem. Phys.* **11**, 5558 (2009).
- <sup>47</sup>P. Mori-Sanchez, A. J. Cohen, and W. T. Yang, *Phys. Rev. A* **85**, 042507 (2012).
- <sup>48</sup>W. T. Yang, Y. K. Zhang, and P. W. Ayers, *Phys. Rev. Lett.* **84**, 5172 (2000).
- <sup>49</sup>Y. K. Zhang and W. T. Yang, *Theor. Chem. Acc.* **103**, 346 (2000).
- <sup>50</sup>P. W. Ayers, *J. Math. Chem.* **43**, 285 (2008).
- <sup>51</sup>J. P. Perdew and M. Levy, *Phys. Rev. B* **56**, 16021 (1997).
- <sup>52</sup>A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy, *Phys. Rev. B* **53**, 3764 (1996).
- <sup>53</sup>A. J. Cohen, P. Mori-Sanchez, and W. T. Yang, *J. Chem. Theory Comput.* **5**, 786 (2009).
- <sup>54</sup>W. Yang, P. W. Ayers, and Q. Wu, *Phys. Rev. Lett.* **92**, 146404 (2004).
- <sup>55</sup>E. Sagvolden and J. Perdew, *Phys. Rev. A* **77**, 012517 (2008).
- <sup>56</sup>W. Kohn, *Phys. Rev. B* **33**, 4331 (1986).
- <sup>57</sup>W. T. Yang and Q. Wu, *Phys. Rev. Lett.* **89**, 143002 (2002).
- <sup>58</sup>Q. Wu and W. T. Yang, *J. Chem. Phys.* **118**, 2498 (2003).
- <sup>59</sup>T. Heaton-Burgess, P. W. Ayers, and W. T. Yang, *Phys. Rev. Lett.* **98**, 036403 (2007).
- <sup>60</sup>P. Mori-Sanchez, A. J. Cohen, and W. T. Yang, *Phys. Rev. Lett.* **102**, 066403 (2009).
- <sup>61</sup>R. G. Parr and L. J. Bartolotti, *J. Phys. Chem.* **87**, 2810 (1983).
- <sup>62</sup>J. F. Janak, *Phys. Rev. B* **18**, 7165 (1978).
- <sup>63</sup>P. Mori-Sanchez, Q. Wu, and W. T. Yang, *J. Chem. Phys.* **123**, 062204 (2005).
- <sup>64</sup>A. Gorling and M. Levy, *Phys. Rev. A* **52**, 4493 (1995).
- <sup>65</sup>M. Gruning, A. Marini, and A. Rubio, *Phys. Rev. B* **74**, 161103 (2006).
- <sup>66</sup>Y. M. Niquet and X. Gonze, *Phys. Rev. B* **70**, 245115 (2004).
- <sup>67</sup>M. Stadele, J. A. Majewski, P. Vogl, and A. Gorling, *Phys. Rev. Lett.* **79**, 2089 (1997).
- <sup>68</sup>U. Salzner and R. Baer, *J. Chem. Phys.* **131**, 231101 (2009).
- <sup>69</sup>R. Baer, E. Livshits, and U. Salzner, *Annu. Rev. Phys. Chem.* **61**, 85 (2010).
- <sup>70</sup>T. Tsuneda, J. W. Song, S. Suzuki, and K. Hirao, *J. Chem. Phys.* **133**, 174101 (2010).
- <sup>71</sup>Q. S. Zhao, R. C. Morrison, and R. G. Parr, *Phys. Rev. A* **50**, 2138 (1994).
- <sup>72</sup>A. Savin, C. J. Umrigar, and X. Gonze, *Chem. Phys. Lett.* **288**, 391 (1998).
- <sup>73</sup>P. W. Ayers, R. C. Morrison, and R. G. Parr, *Mol. Phys.* **103**, 2061 (2005).
- <sup>74</sup>Q. Wu, A. J. Cohen, and W. T. Yang, *Mol. Phys.* **103**, 711 (2005).
- <sup>75</sup>D. J. Tozer and F. De Proft, *J. Phys. Chem. A* **109**, 8923 (2005).
- <sup>76</sup>R. Stowasser and R. Hoffmann, *J. Am. Chem. Soc.* **121**, 3414 (1999).
- <sup>77</sup>E. J. Baerends and O. V. Gritsenko, *J. Phys. Chem. A* **101**, 5383 (1997).
- <sup>78</sup>D. P. Chong, O. V. Gritsenko, and E. J. Baerends, *J. Chem. Phys.* **116**, 1760 (2002).