Chemical Sciences

Review article

Towards a density functional theory of molecular fragments. What is the shape of atoms in molecules?

Hacia una teoría de funcionales de la densidad de fragmentos moleculares. ¿Cuál es la forma de los átomos en las moléculas?

 Victor H. Chávez[1], Adam Wasserman[1,2,*]

[1] Department of Chemistry, Purdue University, West Lafayette, USA
[2] Department of Physics and Astronomy, Purdue University, West Lafayette, USA

Adam Wasserman, Inaugural article for admission as a Corresponding Member to the Academia Colombiana de Ciencias Exactas, Físicas y Naturales

Abstract

In some sense, quantum mechanics solves all the problems in chemistry: The only thing one has to do is solve the Schrödinger equation for the molecules of interest. Unfortunately, the computational cost of solving this equation grows exponentially with the number of electrons and for more than ~100 electrons, it is impossible to solve it with chemical accuracy (~ 2 kcal/mol). The Kohn-Sham (KS) equations of density functional theory (DFT) allow us to reformulate the Schrödinger equation using the electronic probability density as the central variable without having to calculate the Schrödinger wave functions. The cost of solving the Kohn-Sham equations grows only as N^3, where N is the number of electrons, which has led to the immense popularity of DFT in chemistry. Despite this popularity, even the most sophisticated approximations in KS-DFT result in errors that limit the use of methods based exclusively on the electronic density. By using fragment densities (as opposed to total densities) as the main variables, we discuss here how new methods can be developed that scale linearly with N while providing an appealing answer to the subtitle of the article: What is the shape of atoms in molecules?

Keywords: Density functionals; Electronic structure; Chemical reactivity.

Resumen

En cierta forma, la mecánica cuántica da solución a todos los problemas de la química, lo único que hay que hacer es resolver las ecuaciones de Schrödinger para las moléculas de interés. Desafortunadamente, el costo computacional de resolver estas ecuaciones crece exponencialmente con el número de electrones y para más de ~100 electrones resulta imposible resolverlas con precisión química (~2 kcal/mol). Las ecuaciones de Kohn-Sham (KS) de la teoría del funcional de la densidad (*density functional theory,* DFT) permiten reformular las ecuaciones de Schrödinger usando la densidad de probabilidad electrónica como la variable central sin necesidad de calcular las funciones de onda de Schrödinger. El costo de resolver las ecuaciones de Kohn-Sham solo crece como N^3, donde N es el número de electrones, lo que ha llevado a la inmensa popularidad de la DFT en química. A pesar de esta popularidad, incluso las aproximaciones más sofisticadas de las KS-DFT llevan a errores que limitan el uso de métodos basados exclusivamente en la densidad electrónica. En este artículo se discute cómo pueden desarrollarse nuevos métodos que escalen linealmente con N usando densidades de fragmentos como las variables principales en lugar de densidades totales, así como la forma en que estos métodos proveen una respuesta atractiva a la pregunta del subtítulo: ¿cuál es la forma de los átomos en las moléculas?

**Palabras clave:** Funcionales de la densidad; Estructura electrónica; Reactividad química.

Introduction

"What is the shape of atoms in molecules?" is a question without an answer. Atoms in molecules do not have a real, independent existence. When we say that a water molecule has two hydrogen atoms and one oxygen atom, what we mean is that an appropriate supply

of energy can split the molecule into these three atoms. Each of these atoms, when isolated, is spherical, but the ground-state electronic density of the water molecule is not simply the sum of three spheres. However, it is approximately equal to the sum of three spheres (**Figure 1**) and can be written exactly as the sum of three distorted spheres. Thus, the question of the title may be vague, but it is not meaningless. All molecular information (i.e., what makes a molecule different from the simple sum of its isolated constituents) is coded into a set of atomic density distortions.

There is no unique way of defining these atomic density distortions. In fact, there are infinitely many ways in which one can decompose a given molecular density n (**r**) into the sum of atomic-like functions n_α (**r**) (here we use the subscript α to label atoms). For example, Bader's partitioning (**Bader,** 1990) based on the topology of the electron density is a popular choice with merits and limitations that have been amply discussed (**Nalewajski, et al.,** 2000). In spite of the many possible definitions of atomic densities, there is a unique set of densities $\{n_\alpha$ (**r**)$\}$ that sum up to n (**r**) while minimizing the sum of the atomic energies. It is this special set of densities to which we turn our attention in this article.

The density functional theory (DFT) (**Hohenberg & Kohn,** 1964; **Kohn & Sham,** 1965) establishes that any electronic property P of a molecule is a functional of its ground-state electronic density $P = P[n]$. The uniqueness of the set $\{n_\alpha$ (**r**)$\}$ for a given density allows one to understand molecular properties, in principle, as functionals $P = P[\{n_\alpha$ (**r**)$\}]$ of that set. Decades of research in DFT have taught us how the total density n (**r**) can be used as the main variable in molecular calculations, as explained briefly in Section II. Our group is investigating how the atomic densities, as opposed to the total molecular density, may be used as the main variables, which we discuss in Section III. This change of perspective has advantages and disadvantages. The most obvious advantage is a significant lowering of the computational cost of the calculations because instead of having to solve the N-electron Schrödinger equation (a second-order differential equation on $3N$ coupled variables), only a small number of independent equations has to be solved, each for less than N electrons. A second advantage will be explained and illustrated in Section IV: By focusing on atomic densities rather than on total molecular densities, one can fix pervasive errors of density functional approximations and significantly improve the accuracy of certain calculations. A third advantage will be discussed in Section V: The chemical reactivity between two
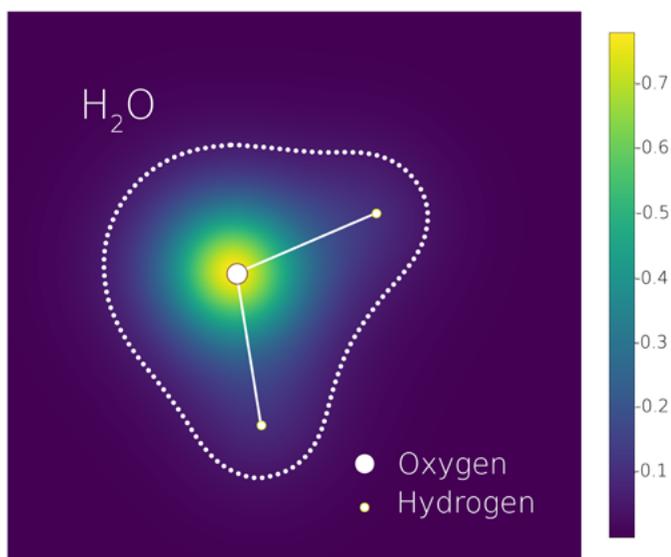
Figure 1. Ground state density of a water molecule on the plane of the three atoms. Calculated using Psi4 (**Smith,** *et al.,* 2018) (**Parrish,** *et al.,* 2017) with CCSD(T)/UGBS. The dotted line indicates a density iso-contour.

atoms or molecular fragments that approach each other is best described in a theory that employs atomic (or fragment) densities as the main variables. On the downside, many of the theorems and techniques that have been explored over the last six decades to describe molecular systems are not directly applicable to the new set of variables, so entirely new methods and approximations need to be developed. A growing community of researchers working under the umbrella of embedding methods (**Jacob & Neugebauer,** 2014; **Lee,** *et al.*, 2019; **Sun & Chan,** 2015; **Nafziger & Wasserman,** 2015; **Niffenegger,** *et al.*, 2019) are pursuing this direction. We highlight here some of the recent developments.

The molecular density as the main variable

In the world of quantum chemistry, one is normally concerned with electronic hamiltonians given within the Born-Oppeheimer approximation by:

$$\hat{H} = \hat{T} + \hat{V}_{ee} + \int d\mathbf{r}\, v(\mathbf{r})\hat{n}(\mathbf{r}) \qquad (1),$$

where $\hat{T}$ is the kinetic energy operator for N electrons, $\hat{V}_{ee}$ the electron-electron repulsion operator, $v(\mathbf{r})$ the attractive potential generated by the nuclei, and $\hat{n}(\mathbf{r})$ the density operator. The expectation value of the latter in the N-electron ground state ψ is the density $n(\mathbf{r}) = \langle\psi|\hat{n}(\mathbf{r})|\psi\rangle$, but an accurate, explicit representation of ψ is unachievable for all but the smallest systems. Ground-breaking work by Kohn, *et al.* (**Hohenberg & Kohn,** 1964; **Kohn & Sham,** 1965) showed that it is possible to construct a theory that completely circumvents the calculation of ψ so that any property, such as the ground-state energy $E = \langle\psi|\hat{H}|\psi\rangle$, can be calculated as a functional of the density. The functional $E[n]$ is normally split into the contribution from the nuclear potential $\int d\mathbf{r}n\,(\mathbf{r})\,v(\mathbf{r})$, and a universal functional $F[n]$ independent of $v(\mathbf{r})$. In Kohn-Sham-DFT (KS-DFT), the most common formulation of DFT, the functional $F[n]$, is decomposed as:

$$F[n] = T_s[n] + \frac{1}{2}\int d\mathbf{r}\int d\mathbf{r}'\,\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\text{XC}}[n] \qquad (2),$$

where $T_s[n]$ is the kinetic energy for an auxiliary system of non-interacting electrons with density $n(\mathbf{r})$, the second term is the Hartree energy, and $E_{XC}[n]$ is the exchange-correlation (XC) energy functional, the only quantity that needs to be approximated in practical calculations. We sometimes combine the Hartree and the XC-contributions into one functional, $E_{HXC}[n]$, whose derivative with respect to the density is a potential, $v_{HXC}(\mathbf{r}) = \delta E_{HXC}[n]/\delta n(\mathbf{r})$. What one solves in practice are the KS equations:

$$\left(-\tfrac{1}{2}\nabla^2 + v_s(\mathbf{r})\right)\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r}) \qquad (3),$$

with $v_s(\mathbf{r}) = v(\mathbf{r}) + v_{HXC}(\mathbf{r})$ leading to $\sum_i |\phi_i(\mathbf{r})|^2 = n(\mathbf{r})$ where the sum goes over the occupied orbitals $\phi_i(\mathbf{r})$. An ample literature exists documenting the successes and failures of different approximations to $E_{XC}[n]$. We point the reader to recent reviews (**Pribram-Jones,** *et al.,* 2015; **Wasserman,** *et al.,* 2017; **Yang,** 2012) and stress that some of the most pervasive failures of approximate KS-DFT appear when stretching chemical bonds (**Dutoi,** *et al.,* 2016; **Makmal,** *et al.,* 2011; **Komsa & Staroverov,** *et al.,* 2016). The large errors observed in these cases encompass both fractional-charge (or delocalization) and fractional-spin (or static correlation) errors (**Cohen,** *et al.,* 2008a; **Cohen,** *et al.,* 2008b; **Mori-Sánchez,** *et al.,* 2008), which are ultimately due to the inability of the approximate XC-functionals to reduce the molecular density to the correct atomic densities (or spin-densities) when bonds are stretched. We illustrate both types of errors below for the local density approximation (LDA), the simplest and earliest approximation for $E_{XC}[n]$ on which the modern ladder of approximations is built (**Perdew,** *et al.,* 2001).

Fractional-charge error

Consider stretching H_2^+. The true ground-state density has left-right symmetry, with $^1/2$ electron on the left atom and $^1/2$ electron on the right atom. The physical state at infinite separation must break this symmetry and produce a neutral hydrogen atom on one side and a bare proton on the other. Both solutions (broken-symmetry and symmetric) should, therefore, have the same energy at infinite separation, but this feature is not achieved by

the LDA or other approximations built upon it. The LDA energy of an H atom with half an electron is much lower than what it should be (half the energy of infinitely stretched H_2^+), which leads to the incorrect binding shown in dashed pink in **figure 2**.

### *Fractional-spin error*

Now consider stretching neutral H_2. The issues of the previous paragraph are no longer a problem because each atom has exactly one electron at infinite separation. However, an analogous problem arises for fractional spins. The true-ground state of H_2 is spin unpolarized and must remain so at very large separations in the absence of environmental perturbations. Yet, an isolated H atom is spin-polarized so the energy of two spin-polarized H atoms should be identical to that of two spin-unpolarized H atoms (each having half-spin up, half-spin down). This condition is again violated by the LDA leading to the significant overestimation of the binding as shown in dashed pink in **figure 3**.

These errors, illustrated here for the two simplest open-shell and closed-shell molecules, are ubiquitous in quantum chemistry. Every time a bond is stretched, as in transition states along chemical reactions, a combination of these errors can creep into and contaminate the DFT calculations. Cancellation between the two errors can occur sometimes (note they have opposite signs) and lead to accidentally accurate results for complex systems, but predicting such cancellations is generally extremely difficult and not something DFT users want to or should rely on. Results from approximate KS-DFT calculations are thus often suspect. A theory that uses atomic densities as the main variables, as opposed to the total molecular density, has the potential to fix such errors. Furthermore, it has the potential to provide information about individual atomic density distortions along with a chemical insight into the reactivity of individual fragments.
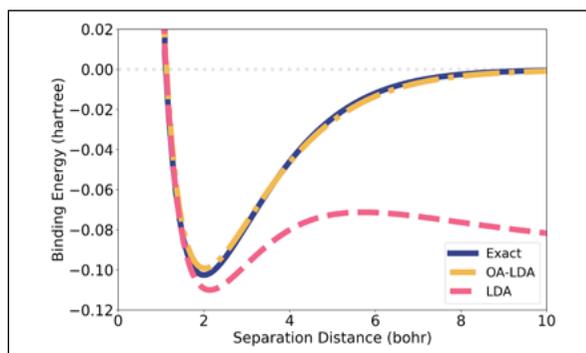
Figure 2. Electronic binding energy of H_2^+. Exact (solid, blue), KS-LDA (dashed, pink), and OA-LDA(dash-dotted, yellow) from (**Nafziger,** 2015), as explained in Sec.IV.
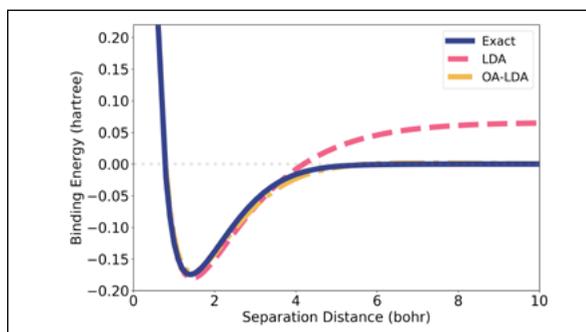
Figure 3. Electronic binding energy of H_2. Exact (solid, blue), KS-LDA (dashed, pink), and OA-LDA (dash-dotted, yellow) from (**Nafziger&Wasserman,** 2015), as explained in Sec.IV.

Atomic densities as the main variables

As mentioned in the Introduction, there is a unique set of atomic densities $\{n_\alpha(\mathbf{r})\}$ that minimizes $E_f \equiv \sum_\alpha E_\alpha[n_\alpha]$ while satisfying the density constraint:

$$\sum_\alpha n_\alpha(\mathbf{r}) = n(\mathbf{r}) \qquad (4).$$

The atomic energies $E_\alpha[n_\alpha]$ in the above definition of E_f are not true ground-state energies but rather given by $E_\alpha[n_\alpha] = F[n_\alpha] + \int d\mathbf{r}\, v_\alpha(\mathbf{r})\, n_\alpha(\mathbf{r})$, where $v_\alpha(\mathbf{r})$ is the α-atomic potential. The constraint of Equation (4) prevents $n_\alpha(\mathbf{r})$ from being the ground-state density of the corresponding $v_\alpha(\mathbf{r})$. However, each of the $n_\alpha(\mathbf{r})$ can be shown to be the ensemble ground-state density of $v_\alpha(\mathbf{r}) + v_p(\mathbf{r})$, where $v_p(\mathbf{r})$ is a unique α-independent potential. More specifically, the partition potential theorem (PPT) (**Cohen & Wasserman**, 2006) establishes the following: If a molecular potential $v(\mathbf{r})$ is decomposed into atomic potentials $\{v_\alpha(\mathbf{r})\}$, i.e. $v(\mathbf{r}) = \sum_\alpha v_\alpha(\mathbf{r})$, then, for a set of fragment occupations $\{n_\alpha\}$ there is a unique local potential $v_p(\mathbf{r})$ such that, when added to the individual $v_\alpha(\mathbf{r})$'s, leads to ensemble ground-state densities $\{n_\alpha\}$ summing up to the correct total density $n(\mathbf{r})$.

Simple Illustration of the PPT

Figure 4 provides the simplest illustration of this theorem. Consider first one electron moving in the 1D-potential $v_1(x)$ (dotted grey line in the middle panel). Its ground-state density, when isolated, is $n_1^{(0)}(x)$ (dotted orange line in the upper panel). Similarly, one electron in $v_1(x)$ has density $n_2^{(0)}(x)$. Now consider two non-interacting electrons in the double-well potential $v(x) = v_1(x) + v_2(x)$ (dashed yellow line in the middle panel). The density of this system, $n(x)$, is not equal to $n^{(0)}(x) = n_1^{(0)}(x) + n_2^{(0)}(x)$, but it is close to it: $n(x) \approx n^{(0)}(x)$, especially if $v_1(x)$ overlaps weakly with $v_2(x)$. The PPT establishes that there is only one potential $v_p(x)$ (purple line in the middle panel) such that, when added separately to $v_1(x)$ (blue line in the lower panel) and $v_2(x)$, leads to ground-state densities $n_1(x)$ (blue line in the upper panel) and $n_2(x)$ that differ from $n_1^{(0)}(x)$ and $n_2^{(0)}(x)$ in just the right way, so that $n_1(x) + n_2(x) = n(x)$. The theorem applies to any number of interacting electrons in 3D and to any number of fragments (**Cohen & Wasserman**, 2006).

The algorithm to calculate $v_p(\mathbf{r})$

Several algorithms have been developed to solve the constrained optimization problem involved in calculating $v_p(\mathbf{r})$ (i.e., minimizing $E_f[n_\alpha]$ under the constraint of Equation (4)). The algorithm described here is perhaps conceptually the simplest:

1. Choose an approximation for $E_{XC}[n]$, solve the KS equations in Equation (3) for the isolated atoms, and find their self-consistent densities $\{n_\alpha^{(0)}(\mathbf{r})\}$ and corresponding KS-potentials $\{v_{s,\alpha}^{(0)}(\mathbf{r})\}$.
2. Build an approximate molecular density as $n^{(0)}(\mathbf{r}) \approx \sum_\alpha n_\alpha^{(0)}(\mathbf{r})$.
3. Invert the KS equations to find the effective KS-potential $v_s^{(0)}(\mathbf{r})$ corresponding to $n^{(0)}(\mathbf{r})$. For the exact density, this potential would be identical to $v(\mathbf{r}) + v_{HXC}(\mathbf{r})$, but for the approximate density it is not. The difference between the two, therefore, can be used as a correction to generate an improved atomic KS potential $v_{s,\alpha}^{(1)}(\mathbf{r}) = v_{s,\alpha}^{(0)}(\mathbf{r}) + v(\mathbf{r}) + v_{HXC}[n^{(0)}](\mathbf{r}) - v_s^{(0)}(\mathbf{r})$.
4. Solve the KS equations for the atoms again with the improved atomic KS potential and repeat until self-consistency is achieved. If convergence is achieved after iterating k times, then the atomic KS potentials $v_{s,\alpha}^{(k)}(\mathbf{r})$ are given by $v_\alpha(\mathbf{r}) + v_p(\mathbf{r}) + v_{HXC}[n_\alpha^{(k-1)}](\mathbf{r})$. The partition potential emerges as an α-independent piece to be added to the α-nuclear potential.

Just as for the model system of **figure 4**, the main feature of $v_p(\mathbf{r})$ in real diatomic molecules is an attractive well in between the nuclei. The presence of this well distorts the density of each isolated atom by pulling it toward the bonding region. There are generally also positive plateaus in $v_p(\mathbf{r})$ due to kinetic effects (**Nafziger & Wasserman**, 2014) (contributions from the functional derivatives of $T_s[n_\alpha]$) and a singularity at the nuclei whose strength is proportional to the value of the density of one atom at the location of the

other (**Nafziger & Wasserman,** 2014), modifying the isolated-atom density of the nuclei according to Kato's cusp condition. This is our partial response to the question without an answer: "The shape of atoms in molecules is determined by $v_p(\mathbf{r})$."

Although we exemplify the use of Partition-DFT with systems composed of two atoms, it must be noted that the framework is not limited to atomic densities. Any meaningful fragmentation can be applied, such as individual molecules in dimers or functional groups. This selective fragmentation not only has conceptual benefits but also computational because it allows regions to be treated with different levels of theory so that the most reactive regions are calculated using higher-level methods, as is the common practice in WFT-in-DFT embedding (**Lee,** *et al.,* 2019). This is consistent with the goal of the theory to lower the computational cost while still accurately describing the relevant properties of a larger system.

Fixing errors of approximate XC-functionals

Now define the partition energy E_p as the difference between the total energy E and the sum of atomic energies E_f. Using a bold $\mathbf{n}(\mathbf{r})$ to denote the set of atomic densities, $\mathbf{n}(\mathbf{r}) \equiv \{n_\alpha(\mathbf{r})\}$,

$$E_p[\mathbf{n}] = E[n] - E_f[\mathbf{n}] \qquad (5)$$

with the KS energy-decomposition of Equation (2), the partition energy is given by:

$$E_p[\mathbf{n}] = T_s^{\text{nad}}[\mathbf{n}] + E_{\text{HXC}}^{\text{nad}}[\mathbf{n}] + V_{\text{ext}}^{\text{nad}}[\mathbf{n}] \qquad (6),$$

where the "nad" superscript is used to indicate nonadditive quantities, i.e., the difference between the total and the sum of the fragments. Thus, $T_s^{\text{nad}}[\mathbf{n}] = T_s[n] - \sum_\alpha T_s[n_\alpha]$, $E_{\text{HXC}}^{\text{nad}}[\mathbf{n}] = E_{\text{HXC}}[n] - \sum_\alpha E_{\text{HXC}}[n_\alpha]$, and the last term stands for the non-additive external
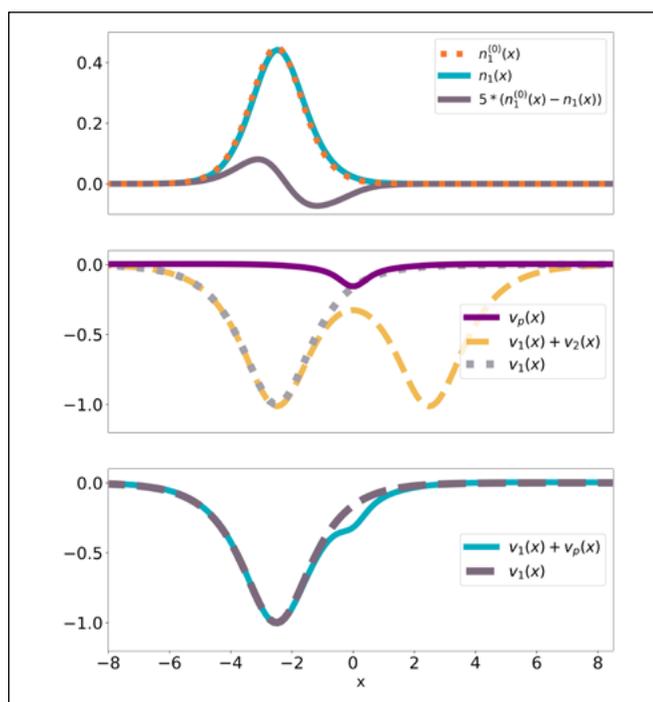
Figure 4. Graphic example of simplest PPT case for 2 non-interacting electrons in 1D potentials of the form $v_{1,2}(x) = \cosh^{-2}(x \pm a)$, with a = 2.5. *Upper panel:* The gray line magnifies the atomic-density distortion $n_1(x) - n_1(x)^0$ by a factor of 5 to highlight what occurs upon formation of the chemical bond: The density of the isolated atom is pulled towards the bonding region. *Middle panel:* Left atomic potential $v_1(x)$ (dotted), total 'molecular potential' (dashed), and partition potential (solid, purple). *Bottom panel*: Comparison of the isolated atomic potential $v_1(x)$ and the effective potential $v_1(x) + v_p(x)$ for which the polarized density $n_1(x)$ is a ground-state density.

(nuclear) energy given by $V_{\text{ext}}^{\text{nad}}[\mathbf{n}] = \int d\mathbf{r}\, n(\mathbf{r})v(\mathbf{r}) - \sum_\alpha \int d\mathbf{r}\, n_\alpha(\mathbf{r})v_\alpha(\mathbf{r})$. We have again combined the non-additive exchange-correlation and Hartree terms of Equation (2) into $E_{\text{HXC}}^{\text{nad}}$. For a given approximation to $E_{XC}[n]$, the algorithm described in the previous section exactly reproduces the results of the corresponding KS calculation including all of its errors. It can be seen (**Elliot, *et al.,*** 2010) that the partition potential is the functional derivative of the partition energy with respect to any of the atomic densities at the minimum. Since $V_{\text{ext}}^{\text{nad}}[\mathbf{n}]$ is known exactly, the key to improving over Kohn-Sham is to propose adequate approximations for the other two components of Equation (6), $T_s^{\text{nad}}[\mathbf{n}]$ and $E_{\text{HXC}}^{\text{nad}}[\mathbf{n}]$. We discuss each of these separately.

Approximating $E_{\text{HXC}}^{\text{nad}}[\mathbf{n}]$

The origin of common errors of approximate XC functionals is well understood especially for homonuclear diatomic molecules (**Cohen, *et al.,*** 2008b) (see discussion above for H_2 and H_2^+). For these and other molecules composed of two fragments, A and B, we have shown how a simple overlap approximation (OA) to $E_{\text{HXC}}^{\text{nad}}[\mathbf{n}]$ can fix both errors simultaneously (**Nafziger & Wasserman,** 2015), a result that no approximate XC functional can achieve with comparable accuracy within KS-DFT. The OA is defined by:

$$E_{\text{HXC}}^{\text{nad},\text{OA}}[\mathbf{n}] = E_{\text{H}}^{\text{nad}}[\mathbf{n}] + S[\mathbf{n}]E_{\text{XC}}^{\text{nad}}[\mathbf{n}] + (1 - S[\mathbf{n}])\Delta E_{\text{H}}^{\text{nad}}[\mathbf{n}] \tag{7},$$

where $S[\mathbf{n}]$ is a measure of the overlap between the two atomic densities:

$$S[\mathbf{n}] = \text{erf}(2\int d\mathbf{r}\, \sqrt{n_A(\mathbf{r})n_B(\mathbf{r})}) \tag{8}.$$

The term $\Delta E_{\text{H}}^{\text{nad}}[\mathbf{n}]$ in Eq.(7) is a correction to the non-additive Hartree contribution, so that both fractional-charge and fractional-spin errors are suppressed as the molecule is stretched (**Nafziger & Wasserman,** 2015). The results shown in yellow in **figure 2** and labeled "OA-LDA" go a step further (**Nafziger,** 2015) and replace $E_{\text{HXC}}^{\text{nad}}[\mathbf{n}]$ in Equation (7) by the non-additive exact-exchange functional canceling completely the self-interaction error and leading to the exact $E_p[\mathbf{n}]$ in this case. The only deviation from the exact binding here is due to the effect that the slightly incorrect LDA fragment densities have on $E_p[\mathbf{n}]$. We stress that the correct energy is obtained here as the molecule is stretched without symmetry breaking: The ground state of H_2^+ retains left-right symmetry and the ground-state of H_2 remains a spin-singlet throughout the entire range of separations. This is a proof-of-principle demonstration that it is possible to use a simple functional of the density for the atoms (not for the molecule) while approximating $E_{\text{HXC}}^{\text{nad}}[\mathbf{n}]$ to fix the underlying errors due to fragmentation. The route is complementary to the efforts of many others to develop sophisticated XC-functionals of the total density (**Hedge,** 2017; **Mardirossian & Head-Gordon,** 2017; **Sun, *et al.,*** 2015; **Yu, *et al.,*** 2016; **Zhang,** 2018).

Approximating $T_s^{\text{nad}}[\mathbf{n}]$

Even with a robust and accurate functional for $E_{\text{HXC}}^{\text{nad}}[\mathbf{n}]$ (based perhaps on future generalizations of the OA), an explicit approximation for $T_s^{\text{nad}}[\mathbf{n}]$ is needed if one wants to reach the goal of linear-scaling calculations. Step 3 of the algorithm described in Section III relies on iterative inversions that make the method computationally impractical (at least, not more efficient than regular KS-DFT). The most obvious way to avoid such inversions is by resorting to orbital-free DFT (OF-DFT) (**Chen, *et al.,*** 2016) and approximate $T_s^{\text{nad}}[\mathbf{n}]$ as an explicit functional of the set of atomic densities. Kinetic-energy functionals are famously difficult to approximate and state-of-the-art functionals are still far from reaching chemical accuracy, but we are looking for approximations to the non-additive part of $T_s[\mathbf{n}]$, which is altogether a different challenge. The non-interacting kinetic energies for the atoms are still calculated exactly via orbitals and we wish to approximate the much smaller $T_s^{\text{nad}}[\mathbf{n}]$. Recent work shows that this approach is promising. For example, writing $T_s^{\text{nad}}[\mathbf{n}] = \int d\mathbf{r}\, n(\mathbf{r})t_s^{\text{nad}}(\mathbf{r})$, an expression for the non-additive kinetic energy density $t_s^{\text{nad}}(\mathbf{r})$ of the type

$$t_s^{\text{nad}}(\mathbf{r}) = Q[\mathbf{n}]t_s^{\text{nad},vW}(\mathbf{r}) + (1 - Q[\mathbf{n}])t_s^{\text{nad},\text{TF}}(\mathbf{r}) \tag{9}$$

has been shown (**Jiang, *et al.,*** 2018) to provide an excellent approximation to the exact T_s^{nad} in covalent σ-bonds. In Equation (9), $t_s^{\text{nad,Vw}}(\mathbf{r})$ is the von Weizsäcker kinetic-energy density (**Weizsäcker,** 1935), which is exact for one-orbital systems, $t_s^{\text{nad,TF}}(\mathbf{r})$ is the Thomas-Fermi kinetic-energy density (**Fermi,** 1927; **Thomas,** 1927), exact for uniform-density systems, and $Q[\mathbf{n}]$ is a switching functional whose role is analogous to that of $S[\mathbf{n}]$ in Equation (7), as described in detail in **Jiang, *et al.,*** 2018. It determines the spatial regions where a one-orbital description should dominate.

Embedding methods provide the most promising route to truly calculate *ab-initio* electronic structures of large, complex, molecular systems (**Yao, *et al.,*** 2019; **Lee, *et al.,*** 2019; **Mi&Pavanello,** 2020). Although Partition-DFT has been applied so far only to model systems (**Elliot, *et al.,*** 2010), diatomic molecules (**Nafziger & Wasserman,** 2014) and small hydrogen-bonded clusters (**Gómez, *et al.,*** 2017; **Gómez, *et al.,*** 2019), an ongoing implementation into the Psi4 package (**Chávez, *et al.,*** 2020) will allow more for more widespread application. An accurate and explicit functional of the fragment densities for $T_s^{\text{nad}}[\mathbf{n}]$ will make Partition-DFT amenable to linear scaling implementations and, thus, applicable to systems of ever-larger complexity.

Towards a quantum theory of chemical reactivity

We glossed over one key feature of the PPT of Section III. The minimization of $E_f[\mathbf{n}]$ is performed under the constraint that the $n_a(\mathbf{r})$ add to $n(\mathbf{r})$ without individual normalization constraints for the atomic densities. The $n_a(\mathbf{r})$ can integrate to fractional numbers of electrons N_a as long as the total density integrates to the correct number of electrons, i.e., $\sum_a N_a = N$. A sensible interpretation of fractional number densities is provided by an ensemble description where the fractional number arises as an ensemble average over integer-number components. One result of the extension of DFT for fractional electron numbers (**Perdew, *et al.,*** 1982) is that for N_a between the integers p_a and p_{a+1}, the minimizing density (for the exact E_{XC}) is given by $n_a(\mathbf{r}) = (1 - \omega_a) n_{pa}(\mathbf{r}) + \omega_a n_{pa+1}(\mathbf{r})$, where $0 \leq \omega_a \leq 1$. The minimization of $E_f[\mathbf{n}]$ is to be performed over the set of the $\{\omega_a\}$ leading to possibly fractional-number densities. One immediate advantage of such fractional densities is that chemical reactivity indices involving derivatives of various properties with respect to the electron number become sharply defined for the atoms (**Geerlings, *et al.,*** 2003; **Geerlings, *et al.,*** 2014). For example, Fukui functions are given directly by

$$f_\alpha^+(\mathbf{r}) = \left.\frac{\partial n_\alpha(\mathbf{r})}{\partial N_\alpha^+}\right|_{v_\alpha} = n_{p_{\alpha+1}}(\mathbf{r}) - n_{p_\alpha}(\mathbf{r}) \tag{10}$$

as normally defined (**Parr & Yang,** 1984), but here the bordering-integer densities include the polarizing effect of the partition potential which accounts for the detailed environment of the atom in the molecule. Similarly, the $\{\omega_a\}$ are those that lead to electronegativity equalization (**Cohen & Wasserman,** 2006; **Parr, *et al.,*** 1978).

Perspective

The approximations of Equations (7) and (9) need to be extended to be applicable to realistic, complex chemical systems. The roads to robust approximations of general applicability and to efficient and accurate linear-scaling algorithms will be long and winding. Therefore, one might wonder whether these roads are worth taking in the first place, especially given that: (a) As computers get more powerful and machine-learning conquers the quantum-chemistry landscape, KS-DFT calculations with sophisticated approximations to E_{XC} will become applicable to an ever-expanding frontier of chemical complexity; (b) several other fragment-based (**Gordon, *et al.,*** 2012) and embedding methods (**Jacob & Neugebauer,** 2014; **Lee, *et al.,*** 2019; **Sun & Chan,** 2016) are enabling multi-million atom and multi-scale calculations where individual fragment densities are of little use, and (c) quantum computing, when finally here, will allow for the direct calculation of many-electron wave functions rendering DFT-based methods obsolete.

Our current take on these three valid concerns is the following: (a) Further improvements of approximate XC functionals will continue via two directions: A non-empirical approach in which more exact constraints will be incorporated, especially perhaps in the framework of Generalized-Kohn-Sham (**Kummel & Kronik,** 2008; **Seidl, *et al.,*** 1996), and an empirical approach exploiting large data sets of chemical information through machine-learning tools (**Seino, *et al.,*** 2018; **Snyder, *et al.,*** 2012). All of these positive developments can be readily incorporated into the framework described in Section III. Furthermore, as mentioned toward the end of Section III, the fragments do not need to be atoms but could be monomers, functional groups, protein backbones, etc., and calculations will greatly benefit from improved approximations to E_{XC}; (b) there is plenty of room in quantum chemistry for more than one type of embedding method. When minimizing the total energy is the only goal, our approach is admittedly not essential. When, however, besides minimizing the energy one is interested in examining the individual fragment density distortions or understanding the reactivity of one fragment in a specific chemical environment, then our approach offers a unique, useful perspective; (c) yes, someday quantum computers will be ready to solve the many-electron Schrödinger equation for large molecules. However, one will always want to understand the results, which involves determining how individual atoms or fragments in the molecule are distorted due to the interactions with neighboring atoms or fragments. The tools described here allow us to accomplish precisely this, regardless of the type of computers employed.

Acknowledgements

Authors' contribution

AW has supervised the writing of the manuscript. VHC was in charge of producing the calculations for the figures and illustrations. All the authors provided critical feedback and helped shape the research and the manuscript.

Conflict of interest

None to declare

References

Bader, R. F. W. (1990). International Series of Monographs on Chemistry. Atoms in Molecules, A Quantum Theory, 22.

Chen, M., Jiang, X. W., Zhuang, H. L., Wang, L. W., Carter, E. A. (2016). Peta-scale orbital-free density functional theory enabled by small-box algorithms. J. Chem. Theory Comput. **12:** 2950-2963.

Cohen, A. J., Mori-Sánchez, P., Yang, W. (2008a). Fractional spins and static correlation error in density functional theory. J. Chem. Phys. **129:** 121104.

Cohen, A. J., Mori-Sánchez, P., Yang, W. (2008b). Insights into current limitations of density functional theory. Science. **321:** 792-794.

Cohen, M. H. & Wasserman, A. (2006). On Hardness and Electronegativity Equalization in Chemical Reactivity Theory. J. Stat. Phys. **125:** 1121-1139.

Dutoi, A. D. & Head-Gordon, M. (2006). Self-interaction error of local density functionals for alkali–halide dissociation. Chem. Phys. Lett. **422** (1-3): 230-233.

Elliott, P., Burke, K., Cohen, M. H., Wasserman, A. (2010). Partition density-functional theory. Phys. Rev. A. **82:** 024501.

Fermi, E. (1927). Statistical method to determine some properties of atoms. Rend. Accad. Naz. Lincei. **6:** 5.

Geerlings, P., De Proft, F., Langenaeker, W. (2003). Conceptual density functional theory. Chem. Rev. **103** (5): 1793-1874.

Geerlings, P., Fias, S., Boisdenghien, Z., De Proft, F. (2014). Conceptual DFT: Chemistry from the linear response function. Chem. Soc. Rev. **43** (14): 4989-5008.

Gómez, S., Nafziger, J., Restrepo, A., Wasserman, A. (2017). Partition-DFT on the water dimer. J. Chem. Phys. **146** (7): 074106.

Gómez, S., Oueis, Y., Restrepo, A., Wasserman, A. (2019). Partition potential for hydrogen bonding in formic acid dimers. Int. J. Quantum Chem. **119** (4): e25814.

Gordon, M. S., Fedorov, D. G., Pruitt, S. R., Slipchenko, L. V. (2012). Fragmentation Methods: A Route to Accurate Calculations on Large Systems. Chem. Rev. **112:** 632-672.

Hegde, G. & Bowen, R. C. (2017). Machine-learned approximations to density functional theory hamiltonians. Sci. Rep. **7:** 42669.

Hohenberg, P. & Kohn, W. (1964). Inhomogeneous electron gas. Phys. Rev. **136:** B864-B871.

Jacob, C. R. & Neugebauer, J. (2014). Subsystem density-functional theory. Wiley Interdisciplinary Reviews-Computational Molecular Science. **4:** 325-362.

Jiang, K., Nafziger, J., Wasserman, A. (2018). Constructing a non-additive noninteracting kinetic energy functional approximation for covalent bonds from exact conditions. The Journal of chemical physics. **149:** 164112.

Kohn, W. & Sham, L. J. (1965). Self-consistent equations including exchange and correlation effects. Phys. Rev. **140:** A1133.

Komsa, D. N. & Staroverov, V. N. (2016). Elimination of spurious fractional charges in dissociating molecules by correcting the shape of approximate Kohn–Sham potentials. J. Chem. Theory Comput. **12** (11): 5361-5366.

Kummel, S. & Kronik, L. (2008). Orbital-dependent density functionals: Theory and applications. Rev. Mod. Phys. **80:** 3-60.

Lee, S. J., Welborn, M., Manby, F. R., Miller III, T. F. (2019). Projection-based wave function-in-DFT embedding. Accounts of Chemical Research. **52:** 1359-1368.

Makmal, A., Kuemmel, S., Kronik, L. (2011). Dissociation of diatomic molecules and the exact-exchange Kohn-Sham potential: The case of LiF. Phys. Rev. A. **83** (6): 062512.

Mardirossian, N. & Head-Gordon, M. (2017). Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals. Mol. Phys. **115:** 2315-2372.

Mi, W. & Pavanello, M. (2019). Nonlocal Subsystem Density Functional Theory. J. Phys. Chem. Lett. **11** (1): 272-279.

Mori-Sánchez, P., Cohen, A. J., Yang, W. (2008). Localization and delocalization errors in density functional theory and implications for band-gap prediction. Phys. Rev. Lett. **100:** 146401.

Nafziger, J. (2015). Partition density functional theory. Ph.D thesis, Purdue University.

Nafziger, J. & Wasserman, A. (2014). Density-Based Partitioning Methods for Ground-State Molecular Calculations. J. Phys. Chem. A. **118:** 7623-7639.

Nafziger, J. & Wasserman, A. (2015). Fragment-based treatment of delocalization and static correlation errors in density-functional theory. J. Chem. Phys. **143:** 234105.

Nalewajski, R. F. & Parr, R. G. (2000). Information theory, atoms in molecules, and molecular similarity. Proc. Natl. Acad. Sci. U.S.A. **97** (16): 8879-8882.

Niffenegger, K., Oueis, Y., Nafziger, J., Wasserman, A. (2019). Density embedding with constrained chemical potential. Mol. Phys. **117** (15-16): 2188-2194.

Parr, R. G., Donnelly, R. A., Levy, M., Palke, W. E. (1978). Electronegativity - density functional viewpoint. J. Chem. Phys. **68:** 3801-3807.

Parr, R. G. & Yang, W. T. (1984). Density functional-approach to the frontier-electron theory of chemical-reactivity. J. Am. Chem. Soc. **106:** 4049-4050.

Parrish, R. M., Burns, L. A., Smith, D. G. A., Simmonett, A. C., DePrince III, A. E., Hohenstein, E. G., Bozkaya U., Sokilov A. Y., Di Remigio R., Richard R. M., Gonthier J. F., James A. M., McAlexander H. R., Kumar A., Saitow M., Wang X., Pritchard B. P., Verma P., Shaefer H. F., Patkowski K., King R. A., Valeev E. F., Evangelista F. A., Turney J. M., Crawford T. D., Sherrill C. D. (2017). Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability. J. Chem. Theory Comput. **13** (7): 3185-3197.

Perdew, J. P., Parr, R. G., Levy, M., Balduz Jr, J. L. (1982). Density-functional theory for fractional particle number: Derivative discontinuities of the energy. Phys. Rev. Lett. **49:** 1691-1694.

Perdew, J. P. & Schmidt, K. (2001). Jacob's ladder of density functional approximations for the exchange-correlation energy. AIP Conf. Proc. **577:** 1-20.

Pribram-Jones, A., Gross, D. A., Burke, K. (2015). DFT: A Theory Full of Holes? Annu. Rev. Phys. Chem. **66:** 283-304.

Seidl, A., Gorling, A., Vogl, P., Majewski, J. A., Levy, M. (1996). Generalized Kohn-Sham schemes and the band-gap problem. Phys. Rev. B. **53:** 3764-3774.

Seino, J., Kageyama, R., Fujinami, M., Ikabata, Y., Nakai, H. (2018). Semi-local machine-learned kinetic energy density functional with third-order gradients of electron density. J. Chem. Phys. **148:** 241705.

Smith, D.G., Burns, L.A., Sirianni, D.A., Nascimento, D.R., Kumar, A., James, A.M., Schriber, J.B., Zhang, T., Zhang, B., Abbott, A.S., Berquist, E.J. (2018). Psi4numpy: An interactive quantum chemistry programming environment for reference implementations and rapid development. Journal of chemical theory and computation. **14:** 3504-3511.

Snyder, J. C., Rupp, M., Hansen, K., Muller, K. R., Burke, K. (2012). Finding density functionals with machine learning. Phys. Rev. Lett. **108:** 253002.

Sun, J., Ruzsinszky, A., Perdew, J. P. (2015). Strongly constrained and appropriately normed semi-local density functional. Phys. Rev. Lett. **115:** 036402.

Sun, Q. N. & Chan, G. K. L. (2016). Quantum embedding theories. Acc. Chem. Res. **49:** 2705-2712.

Thomas, L. H. (1927). The calculation of atomic fields. P. Camb. Philos. Soc. **23:** 542-548.

Chávez V.H., Shi Y., Oueis Y., Wasserman. A. (2020). Basis set implementation of partition density functional theory. In progress.

Wasserman, A., Nafziger, J., Jiang, K. L., Kim, M. C., Sim, E., Burke, K. (2017). The importance of being inconsistent. Annu. Rev. Phys. Chem. **68:** 555-581.

Weizsäcker, C. F. von. (1935). Zur theorie der kernmassen. Zeitschrift fur Physik A Hadrons and Nuclei. **96:** 431-458.

Yang, W., Cohen, A. J., Mori-Sánchez, P. (2012). Derivative discontinuity, bandgap and lowest unoccupied molecular orbital in density functional theory. J. Chem. Phys. **136:** 204111.

Yao, Y., Shushkov, P., Miller, T. F., Giapis, K. P. (2019). Direct dioxygen evolution in collisions of carbon dioxide with surfaces. Nat. Commun. **10** (1): 2294.

Yu, H. Y. S., Li, S. H. L., Truhlar, D. G. (2016). Perspective: Kohn-sham density functional theory descending a staircase. J. Chem. Phys. **145:** 130901.

Zhang, Y., Kitchaev, D.A., Yang, J., Chen, T., Dacek, S.T., Sarmiento-Pérez, R.A., Márques, M.A., Peng, H., Ceder, G., Perdew, J.P., Sun, J. (2018). Efficient first-principles prediction of solid stability: Towards chemical accuracy. NPJ Comput. Mater. **4:** 9.