Cite this: DOI: 00.0000/xxxxxxxxxx

Localizing Electron Density Errors in Density Functional Theory[†]

Rubén Laplaza,^a Victor Polo,^b and Julia Contreras-García^{*c}

Received Date

Accepted Date

DOI: 00.0000/xxxxxxxxxx

The accuracy of different Density Functional Approximations is assessed through the use of Quantum Chemical Topology on molecular electronic densities. In particular, three simple yet ever-important systems are studied: N₂, CO and ethane. Our results exemplify how real-space descriptors can help understand the sources of error in Density Functional Theory, avoiding unwanted error compensation present in simplified statistical metrics. Errors in “well” built functionals are shown to be concentrated in chemically meaningful regions of space, and hence they are predictable. Conversely, strongly parametrized functionals show isotropic errors that cannot be traced back to chemical transferable units. Moreover, we will show that energetic corrections are mapped back into improvements in the density in chemically meaningful regions. These results point at the relevance of real-space perspectives when parametrizing or relating energy and density errors.

Density Functional Approximations (DFAs) to the exact exchange-correlation functional ($E_{xc}[\rho]$) are usually ordered along Jacob's Ladder¹ in which higher rungs generally include more sources of information. An increase in sophistication, however, does not guarantee an increase in performance. While an increase in the number of sources of information does necessarily increase the flexibility of DFAs, allowing them to reproduce features that simpler models could not, improvement can be insufflated into DFAs through extensive parametrization and fitting. This trend poses two troublesome questions: first, one must determine which properties are the ones a fitted DFA should aim to reproduce; and subsequently, what systems should be used as benchmark for those properties. To reduce the number of variables these questions present, a viable strategy is to simply reduce the scope of a given DFA to a certain set of systems and properties. Alternatively, the creation of vast data sets of chemical information, or the careful selection of few yet representative features, can soundly lay the foundation of parametrization strategies. Both a matter of philosophy and convenience, the amount of such parameters varies wildly among DFAs. Two general trends with respect to parametrization seem to persist: either trying to avoid it at all costs, continuously searching for strict physical constraints

that can be mathematically transferred to $E_{xc}[\rho]$,² or embracing it as a necessary yet fruitful tool in the development of accurate DFAs with a somewhat limited scope. In all cases, the energy of the system remains the most relevant observable in terms of parametrization, evaluation, and overall usage, be it in absolute or in relative terms.

Attention has been recently drawn by different authors to the quality of $\rho(r)$ in different DFAs.^{3–7} Given the exact expression of $E_{xc}[\rho]$, the energy of a many-electron system is minimal for the exact density distribution. However, within any DFA, the relationship between energies and densities remains obscure, and uncontrolled effects can lead to precise results in both features, or the improvement of one at the cost of the other. Alas, in DFT, the right answer might not be for the right reason yet, as recently suggested.⁸ It has been proposed that excessive parametrization can lead to good thermochemical behavior precisely at the cost of a good description of $\rho(r)$ and its derived scalar fields. After all, a functional fitted for the energy may be inaccurate for the derivatives of the energy. However, no conclusion is to be found indisputable as long as a basic question remains open: how, exactly, should we evaluate the quality of these fields? In fact, $\rho(r)$ is well known for being rather robust with respect to the calculation level,⁹ meaning that differences are generally small in relative terms.¹⁰ Furthermore, its value is higher in the proximity of the nuclei than in the rest of the molecule by several orders of magnitude, as shown in Figure 1a, and hence the absolute weight of the same relative error quickly explodes for core electrons, which are not necessarily chemically relevant. Hence, normalized metrics or arbitrary partitions can forfeit chemically relevant information. Furthermore, it has been suggested that only cer-

^a Sorbonne Université, CNRS, Laboratoire de Chimie Théorique, LCT, F. 75005 Paris, France and Departamento de Química Física and Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, E. 50009, Zaragoza, Spain

^b Departamento de Química Física and Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, E. 50009, Zaragoza, Spain

^c Sorbonne Université, CNRS, Laboratoire de Chimie Théorique, LCT, F. 75005 Paris, France; E-mail: contreras@lct.jussieu.fr

[†] Electronic Supplementary Information (ESI) available: See DOI: 10.1039/cXCP00000x/

tain subspaces of $\rho(r)$ are relevant in DFT.^{11,12} On the reverse, different energy-based modifications of DFAs, such as corrections for self-interaction errors, have been shown to affect the density in a similar way in localized regions.^{13,14} Considering this, there is a possibility – and hope – that energy-based parametrization strategies indirectly improve the density in specially meaningful regions of space.

(a) $\rho(r)$, a.u.

(b) $v_{ELF}(r)$

(c) $v_{LOL}(r)$

Fig. 1 Scaled representation of several scalar fields (see definition in the main text) in the σ_v plane of the CO molecule. The electron density in 1a has been truncated at 5 a.u.

From the chemical point of view, it makes perfect sense to ex-

pect that the error enclosed within certain localized regions of space is far more meaningful than others.^{15,16} In this direction, Quantum Chemical Topology (QCT)¹⁷ provides tools for the decomposition of chemical systems in regions through the analysis of scalar fields. It should be noted that these scalar fields are also ingredients that lie in the core of many DFAs. Such is the case for the Localized Orbital Locator (LOL)^{18,19}, the Electron Localization Function (ELF)^{20,21} or the Reduced Density Gradient (RDG)^{22–24}. For instance, both the ELF and the LOL clearly identify and separate core regions from bonds and lone pairs (Figures 1b and 1c). Since they are constructed in terms of derivatives of $\rho(r)$ and/or Kohn-Sham (KS) orbitals, all of these fields depend on the description of the density given by a certain DFA, and any subsequent partition scheme will equally be representative of this description. Our proposal is that by evaluating the quality of $\rho(r)$ and its derived scalar fields with topological methods an increased amount of insight can be extracted about the performance of different DFAs.

1 Computational methods

Following current studies in the field, the all electron quadruple- ξ aug-cc-pCVQZ basis set^{25,26} was used in all calculations for the N₂, CO and ethane molecules, boasting both a set of diffuse functions and a set of tight functions for a correct description of the system far away and close to nuclei. Reference scalar fields were calculated at the all-electron CCSD(T) level using the PSI-4 program version 1.1.²⁷ All calculations and manipulations of the scalar fields were performed using the Multiwfn program version 3.5,²⁸ in which scalar field data was consistently treated with a three-dimensional grid with a 0.026459 Å spacing. Additional details can be found in the ESI†. DFT, Hartree-Fock (HF) and Møller-Plesset second order perturbation theory (MP2) calculations were performed using the Gaussian09 rev.D01 suite²⁹, and therefore all the nomenclature and definitions correspond to those within this software. A total of 59 DFAs, as defined and available in the code, were considered across this work. 27 of them are pure DFAs, and 30 are hybrid DFAs including a percentage of exact exchange (HFX), and 2 are double-hybrids including second order perturbation theory contributions. Core orbitals were frozen to perform the perturbation theory calculations, yet the frozen core (FC) denomination will be omitted for simplicity throughout the text. A detailed list is available in the ESI†.

The systems that have been thoroughly studied, N₂, CO and ethane, are of utmost importance in chemistry. N₂ and CO form an educational duo³⁰ and sample different levels of polarity, while ethane represents a simpler single bond. In some cases, the CCSD(T)/aug-cc-pCVQZ optimized equilibrium geometries have been used, with equilibrium bond lengths (r_e) of 1.0984 Å and 1.1294 Å for the N₂ and CO molecules, and 1.5223 Å for the C–C bond in ethane. Whenever not explicitly mentioned, molecular geometries have been tightly optimized according to the method. All data are presented in the ESI†(Tables S4-S11).

2 Results and discussion

2.1 Spatial localization of density errors

Errors in DFT can be categorized according to their origin.¹⁰ The majority of the errors arise, naturally, from the inexactitude of the exchange-correlation term in the DFA. These errors have been called functional-based errors, and can be fixed by improving a given DFA. However, a given DFA may be very good for energy *evaluation*, but not at providing the correct $\rho(r)$ in a KS-DFT self-consistent procedure. This gives rise to what have been termed density-based errors; errors that may be corrected by providing an accurate $\rho(r)$ onto which evaluate a given DFA.

Ultimately, density-based errors stem from the DFA: an unambiguous, univocal division is not available. This is so because the energies given by two DFAs with their own self-consistent densities are separated by two different paths, each one comprising a density-based contribution (the change in the energy due to the approximate density) and a functional-based contribution (the change in the energy due to the DFA). In general, this distinction suffices to ascertain whether a system can be considered normal (i.e., prone to functional errors, robust with respect to the density) or not.

In any case, it is undeniable that the way in which we discuss errors in DFAs tends to be chemically intuitive. Terms like overbinding and overdelocalizing, used to describe failures in DFT, are easily conceptually mapped into $\rho(r)$: for a given DFA, slightly increasing the density in a covalent bond at the equilibrium distance should allegedly overbind with respect to the self-consistent density of the method. Analogously, overdelocalization can be thought of as the tendency of most DFAs to spread the electron density in an homogenized way from the nuclear cusps onto the low density regions. Covalent bonds are local entities with a certain degree of charge accumulation, and therefore they are affected strongly by delocalization errors.³¹ As such, a correct description of $\rho(r)$ in covalent bonds is highly important for any general purpose DFA.

Let us define a simple function to evaluate the quality of $\rho(r)$. As a reference, a CCSD(T) level density can be used. It must be noted that the use of the CCSD(T) density as a reference in this work merely aims to provide an example of a balanced charge distribution (e.g. the correct dipole moment for CO), belonging to an accurate wave function. The electronic density difference $\rho_{diff}(r)$ with respect to the reference can be calculated as (Equation 1):

$$\rho_{diff}(r) = \rho_{CCSD(T)}(r) - \rho(r) \quad (1)$$

which can be integrated to obtain a simple global descriptor Λ_{diff} (Equation 2),

$$\Lambda_{diff} = \int |\rho_{diff}(r)| dr \quad (2)$$

that formally corresponds to a number of electrons. The role of this integral is akin to the root mean square deviation between the two distributions, but converges with the number of grid points. This is easily verified whenever $\int \rho_{diff}(r) dr = 0$.

On the other hand, defining bonding regions is feasible through the analysis of the changes in the kinetic energy density, $\tau(r)$. Partitioning the system into local subspaces allows for the isolation of

the problem. As a prospective tool, the Localized Orbital Locator (LOL) is one of the simplest indicators that can identify covalent bonds. It is defined as the dimensionless bounded variable $v(r)$ (Equation 3),

$$v_{LOL}(r) = \frac{t_{LOL}(r)}{1 + t_{LOL}(r)} \quad (3)$$

where $t_{LOL}(r)$ is simply the ratio (Equation 4)

$$t_{LOL}(r) = \tau_{TF}(r)/\tau(r) \quad (4)$$

between the Thomas-Fermi kinetic energy, $\tau_{TF}(r)$ (Equation 5),

$$\tau_{TF}(r) = 3/10(3\pi^2)^{2/3}\rho(r)^{5/3} = C_F\rho(r)^{5/3} \quad (5)$$

and the positive definite kinetic energy density given by the occupied KS orbitals ϕ_i , $\tau(r)$ (Equation 6),

$$\tau(r) = \frac{1}{2} \sum_i^N |\nabla \phi_i(r)|^2 \quad (6)$$

As per the Lorentzian transformation that bounds $v_{LOL}(r)$, an arbitrary isovalue can be chosen to delimit regions around maxima in which the ratio given by $t_{LOL}(r)$ has an upper limit. For instance, $v_{LOL}(r) = 0.5$ is reasonable given the straightforward interpretation of $t_{LOL}(r) = 1$: at higher (lower) $v(r)$, electrons have less (more) kinetic energy than in the Thomas-Fermi model, and therefore are more (less) localized.³² In Figure 2, and to a lesser extent in Figure 1c, it can be clearly seen that the bonding region is approximately identified even by this simple function.

As an example, let us briefly examine the N_2 molecule at the fixed reference geometry. It is readily seen in Figure 3 that, for the simplest DFAs and HF, the sign of $\rho_{diff}(r)$ diverges acutely in the regions enclosed by the $v(r)_{LOL} = 0.5$ isoline: the error in $\rho(r)$ is chemically localized, and where, how and why this error localization takes place has far more significance than its total value over the whole system. Noticeably, $\rho_{diff}(r)$ tends to have its maxima and/or minima close to nuclear positions. However, the relative magnitude of $\rho_{diff}(r)$ with respect to $\rho_{CCSD(T)}(r)$ is of the order of 1% for core electrons, while it can be up to 10–20% in bonding regions. This immediately highlights the difficulties of using statistical descriptors.

Within the triple bond, the relative quality of the density with respect to the reference improves as we move towards higher rungs of DFAs. In this case, LDA, PBE and TPSSh present increasing quality while belonging to a similar trend (Figures 3 and 5), in which $\rho(r)$ is generally higher than the reference in the core region and spreads in a relatively smooth way compared to HF. The vast improvement in $\rho(r)$ shown by the hybrid functional is to be expected, as generally Hartree-Fock and GGAs manifest opposite trends, and thus should somehow compensate each other when combined. M062X, on the other hand, exhibits a different error distribution over the plane. This simple example already captures the problems of Λ_{diff} . Methods with identical Λ_{diff} can present differences in terms of the physics behind them. Some selected $\rho_{diff}(r)$ distributions with their respective Λ_{diff} are shown in Figure 4 in order to showcase this point. More examples can be found in the ESI† (Figures S1-S3). On the contrary, a QCT-based

Fig. 2 N₂, CO and ethane C₂H₆ at their CCSD(T) equilibrium geometries with $v_{LOL}(r) = 0.5$ isosurfaces. Relevant bonding domains are colored green, while lone pairs and C–H domains are presented in red. Hydrogen, carbon, nitrogen and oxygen atoms are depicted as white, grey, blue and red balls-and-sticks respectively.

Fig. 3 Electronic density difference ($\rho_{diff}(r)$) maps on the σ_v planes of the N₂ molecule at r_e . $v_{LOL}(r) = 0.5$ isolines are shown in black. Methods are detailed in the top left corners, left to right: Hartree-Fock, SVWN3, PBEPBE, M062X and TPSSh.

analysis is quantitative, exhaustive, and offers straightforward interpretation.

The ELF provides a similar partitioning and insight as the LOL, but has a most established background towards chemical interpretation. Integrating the electron density over ELF basins (Ω_i), regions of space surrounding maxima in zero-flux surfaces, has been used to understand bonding patterns, reactivity and recover Lewis entities with great success.^{33–36} The descriptor $\eta_{ELF}(r)$ is obtained by the following Lorentzian transformation (Equation 7):

$$\eta_{ELF}(r) = \frac{1}{1 + \chi_{ELF}^2(r)} \quad (7)$$

where $\chi_{ELF}(r)$, the ELF kernel, which is remarkably relevant in DFA development, is now the ratio (Equation 8)

$$\chi_{ELF}(r) = \tau_p(r) / \tau_{TF}(r) \quad (8)$$

between the Pauli kinetic energy density $\tau_p(r)$ from the KS orbitals (Equation 9),

$$\tau_p = \tau(r) - \frac{|\nabla\rho(r)|^2}{8\rho(r)} = \frac{1}{2} \sum_i^N |\nabla\phi_i(r)|^2 - \frac{|\nabla\rho(r)|^2}{8\rho(r)} \quad (9)$$

and the Thomas-Fermi kinetic energy $\tau_{TF}(r)$ (cf. Equation 5).

$\eta_{ELF}(r)$ is hence a three-dimensional continuous and differentiable scalar field. It partitions molecular systems into chemical entities (often called basins) divided by zero-flux surfaces in which all points satisfy $\mathbf{n} \cdot \nabla\eta_{ELF}(r_s) = 0$, with $\mathbf{n}$ being the normal vector to the surface at the point r_s . Lewis entities as atomic cores, bonds and lone pairs can be defined by such partitioning,

Fig. 4 Electronic density difference ($\rho_{diff}(r)$) maps on the σ_v planes of the CO molecule at r_e . Methods are detailed in the top left corners, left to right and top to bottom: BHandH, M052X, wPBEhVWN and M11. Their respective values of Λ_{diff} are: 0.1165, 0.1168, 0.1498 and 0.1503.

Fig. 5 Electronic density difference ($\rho_{diff}(r)$) maps on the σ_v planes of the ethane C_2H_6 molecule at r_e . Methods are detailed in the top left corners, left to right and top to bottom: HF, SVWN3, PBEPBE and M062X. Their respective values of Λ_{diff} are: 7.6555, 7.6672, 7.5222 and 7.5271.

as seen in Figures 1b and 6. The quality of the partitioning depends, as $\rho(r)$ does, on the Kohn-Sham orbitals used to build the kinetic energy density, $\tau(r)$ (cf. Equations 7-9).

In the following sections, the framework of the ELF topological analysis will be used to critically assess differences in the densities provided by different DFAs in chemically relevant regions, and analyze the relative importance and physical origin of the divergences. In particular, we will focus on the ELF basins associated with covalent bonds, Ω_b , shown in a truncated form as $\eta_{ELF}(r) = 0.75$ green isosurfaces in Figure 6 for the different systems.

2.2 Analysis of bond descriptors for simple molecules

For the CO molecule at the r_e geometry, Λ_{diff} was calculated for all the DFA set. The use of the CCSD(T) density as a reference is supported by the calculated dipole moment of 0.1169 D, within the uncertainty of the experimental value of 0.112 ± 0.005 D.³⁷ As shown in Figure 7, Λ_{diff} , for a set of functionals, is not correlated with the *RMSD* of absolute or atomization energies (ΔE_{atom}) with respect to CCSD(T), nor with a better dipole moment, all calculated at r_e . Furthermore, there is no correlation between Λ_{diff} values and the accuracy of the electron population of the triple bond, Ω_b , thus proving that it is a measurement that is highly affected by core regions. More generally, this signals that the errors in the electron density cannot be related to the errors in properties all alone. It should be noted that although the frozen geometry may bias these results, Λ_{diff} and similar descriptors can only bear some significance when comparing densities at equal geometries, in which relatively strong electronic forces may be present for some DFAs.

The observed results are replicated in N_2 and ethane, presented in the ESI†. Neither the errors in the absolute energies nor the

errors in atomization energies correlate with Λ_{diff} . Consequently, the minimization of Λ_{diff} as a DFA parametrization criteria does not seem to be well founded.

Thus, Λ_{diff} is not a valid indicator for quantitatively assessing the quality of the self-consistent density of a DFA. A single energetic criteria may not be sufficient, either. Hence, finding meaningful indices that measure DFA failures in terms of the density is of interest. Focusing only on the electronic population of the bonds given by Ω_b , the disynaptic bonding basin, reveals explicitly the effect of different DFAs in the electronic density (Figure 8). Moreover, it constitutes a transferable index. While the integrated electron density of Ω_b can be computed and compared at a fixed geometry, in this case geometries can be optimized tightly for every method. Note that attractors and basins have been merged accordingly in order to retrieve the full triple bond population, as not all DFAs provide isotopological ELF profiles at the chosen geometry. A selection of results are presented in Table 1 for the N_2 molecule, while the complete set can be found in the ESI† (Tables S6 and S7).

Worth noting, as per the inter-method robustness of $\rho(r)$, the change in geometry from r_e to the minima of the potential energy surface for each DFA may switch a method from an “overpopulated” basin to an “underpopulated” one with respect to the reference, better reflecting the character of the DFA. Conveniently, other integrated topological descriptors, such as Bader charges, are equally comparable at different geometries. In other words, integrated topological quantities provide a transferable and adequate depiction of the behavior of a given DFA.

From an electrostatic point of view, higher charge concentrations in bonding domains should lead to stronger bonds. However, the electronic population of Ω_b does not strongly correlate with the atomization energy, ΔE_{atom} , in the $C\equiv O$ or $N\equiv N$ bonds (Figures 8B and 9B). Neither it does with the Bond Dissociation Energy (BDE) of the C–C bond in ethane (Figure 10B). It should be noted that strong trends may be observed in certain hand-picked subsets (i.e. DFA families with the same correlation term), but this does not provide in any way a rigorous connection between the bonding density and the atomization energy that can be used as a reference. Other properties associated with the intrinsic bond strength are the equilibrium bond lengths (Figures 8A, 9A and 10A) and harmonic frequencies (Figures 8C, 9C and 10C). Volume and population of Ω_b are not inter-correlated (Figures 8D, 9D and 10D), which signals that the integrated electron density of the basin is related to the quality of both $\rho(r)$ and $\tau(r)$, as a larger volume does not imply a higher integrated charge. The complete dataset is provided in the ESI†.

Contradicting information can be drawn from different descriptors. As it can be seen in Table 1 and Figures 8,9 and 10, HF leads to overpopulated bonding basins with respect to the reference, partly because of the lack of electron correlation and partly due to the massive volume of Ω_b (i.e., an excessive lowering of $\tau(r)$ in the bonding region), arising from very localized orbitals. Accordingly, in both N_2 and CO the triple bond is shortened and the associated harmonic frequency is overestimated: all this criteria point to an increasingly strong bond. However, atomization energies, and the Bond Dissociation Energy (BDE) of the C–C bond

Fig. 6 N_2 , CO and ethane C_2H_6 at their CCSD(T) equilibrium geometries with $\eta_{ELF}(r) = 0.75$ isosurfaces. Relevant bonding domains are colored green, while lone pairs are presented in red. Basins corresponding to C–H bonds are not shown. Hydrogen, carbon, nitrogen and oxygen atoms are depicted as white, grey, blue and red balls-and-sticks respectively.

Fig. 7 Correlation between Λ_{diff} and the relative error with respect to the CCSD(T) reference in several properties for the CO molecule at r_e for the DFA set. Least-squares fits are drawn as dashed black lines with their Pearson correlation coefficients (R) shown top left of each plot.

Table 1 Descriptors for selected methods for the respective equilibrium geometries of N_2 . Population of Ω_b in electrons, volume of Ω_b in $\text{\AA}^3$, equilibrium distances (d_e) in $\text{\AA}$, harmonic frequencies (ω_h) in cm^{-1} and atomization energies (ΔE_{atom}) in kcal/mol.

Method	Pop. Ω_b	Vol. Ω_b	d_e	ω_h	ΔE_{atom}
CCSD(T)	3.4074	21.26	1.0983	2346.79	224.61
MP2	3.2911	15.28	1.1105	2201.86	236.53
HF	3.8508	38.26	1.0655	2729.25	117.62
SVWN	3.361	15.82	1.0937	2405.14	274.85
PBEPBE	3.3562	18.46	1.1022	2350.65	243.93
G96LYP	3.302	16.04	1.1008	2340.92	235.9
TPSSh	3.4568	21.46	1.0943	2418.89	222.21
B3LYP	3.4537	20.39	1.0901	2448.19	230.07
M06HF	3.4571	28.27	1.0854	2512.34	227.42
BMK	3.5194	22.83	1.0894	2456.13	229.55
mPW2PLYP	3.4361	20.29	1.0942	2383.69	228.95

in the case of ethane, are very low at the HF level (Figures 8B, 9B and 10B). Correlations between the populations of Ω_b and these features is not strong enough to draw rigorous conclusions, yet all Pearson correlation coefficients are significantly higher than in the case of Λ_{diff} . Altogether, this highlights the decoupling between the quality of $\rho(r)$ and relative energetic measures.

It is reasonable to expect such bond shortening, since a higher negative charge accumulation in the bonding region ought to better alleviate the repulsive electrostatic forces between the positive nuclei. The case of ethane is less straightforward, since density redistributions and associated bond shortening effects are shared between all covalent bonds in the molecule. HF does not induce a significant shortening of the C–C bond in ethane (Figure 10A) but clearly shortens C–H bonds with respect to all the other methods. Pure local DFAs like SVWN present exactly the opposite, with longer C–H bonds (see ESI† for an in-depth discussion).

In general, DFAs from the lowest rungs tend to be opposite to HF theory, concentrating less electron density in covalent bonding regions. This can be appreciated clearly in the distribution maps of $\rho_{diff}(r)$ as shown in the previous section (Figures 3-5). In general, hybrid and meta-GGA DFAs better reproduce the CCSD(T) results. In DFT, the correct description of the bonding region is given by a delicate balance between exchange interactions and Coulombic repulsions. The mismatch between the two terms, at the origin of self-interaction error, has a similar effect on the density as the pair-correlation effects introduced by MP2, pulling density from localized bonding regions towards non-bonded space.³⁸ Within Ω_b , self-interaction mimics instantaneous correlation to some extent.

In the case of N_2 , the MP2 density fails to reproduce the CCSD(T) integrated bonding density. MP2 can be thought to overshoot in this regard, and is shown to be quite far away from the CCSD(T) result in most properties (Figures 8). However, in the case of CO, MP2 reproduces the CCSD(T) density and kinetic energy density quite formidably, giving a population of Ω_b within 0.002 electrons of the reference. Unsurprisingly, MP2 is nearly identical to CCSD(T) in terms of bond length and harmonic frequency, and reasonably close in dipole moment (0.1367 D) in this case (Figure 9).

All in all, when correlation is included from a wavefunction method (self-interaction free), a consistent relationship can be

expected between errors in the density and in the energy. In the CCSD(T) calculations, the effect of double excitations as well as perturbative triples is equally signed and similar in value for **all** systems. Thus, the CCSD(T) reference remains equally appropriate while the MP2 performance changes, in spite of the similarities between N_2 and CO, from a clear overshooting in the former to a very accurate result in the latter.

In all cases the CCSD(T) result lies within the cloud of DFA results (Figures 8, 9 and 10) and roughly lies at separation between pure and hybrid DFAs. For harmonic frequencies, atomization energies and bond lengths, GGAs, meta-GGAs and low-exact exchange functionals tend to fall close to the MP2-to-CCSD(T) level. The success of DFT in any of these features and in bonding densities is remarkable. CCSD(T), as expected, gives harmonic frequencies, distances and atomization energies within 1% of the experimental values. DFAs surrounding the CCSD(T) result are therefore very accurate at a relatively low cost. However, it is also evident that some hybrid DFAs, specially those with a high percentage of exact exchange, tend to deviate towards the HF results.

The pure-hybrid distinction is blurred in the case of ethane (Figure 10). Certain DFAs overlocalize C–C bonds, others favor localizing C–H bonds in their stead. As previously discussed, even the limit case of HF manifests more starkly in C–H bonds. This exemplifies how increasingly polyatomic molecules may achieve, through error compensation, accurate results in DFT.

As the electron density in Ω_b is apparently related to the amount of exact exchange in the DFA formulation, a consistent series of DFAs with increasing HFX was created. Two simple one parameter hybrid DFAs (one based on the well-known B3LYP which we shall designate BX-LYP, and one based in a simple LSDA, which we shall call SX-VWN3) have been used with HFX percentages ranging from 10% to 90%. A systematic nomenclature system will be used to name the resulting DFAs, with B10%LYP designating the 10% HFX BX-LYP variant and so on. Details of the procedure can be found in the ESI†.

As shown in Figure 11, more HFX linearly leads to a higher population on the bond. Analogously, in the rest of the properties previously examined, it appears that the inclusion of larger amounts of HFX contributes to displace the results towards the HF

Fig. 8 Correlation between the population of Ω_b and several descriptors for the N_2 molecule as calculated by the DFA set, HF, MP2 and CCSD(T). **A)** Bond length (Å) **B)** ΔE_{atom} (kcal/mol) **C)** Harmonic frequencies (cm^{-1}) **D)** Ω_b volumes ($\text{\AA}^3$). Pure functionals are colored blue, hybrid and double hybrid functionals are colored green, and wavefunction methods are colored red. Least-squares fits to the DFA set data are drawn as dashed black lines with their Pearson correlation coefficients shown in each plot.

Fig. 9 Correlation between the population of Ω_b and several descriptors for the CO molecule as calculated by the DFA set, HF, MP2 and CCSD(T). **A)** Bond length (Å) **B)** ΔE_{atom} (kcal/mol) **C)** Harmonic frequencies (cm^{-1}) **D)** Ω_b volumes ($\text{\AA}^3$). Pure functionals are colored blue, hybrid and double hybrid functionals are colored green, and wavefunction methods are colored red. Least-squares fits to the DFA set data are drawn as dashed black lines with their Pearson correlation coefficients shown in each plot.

Fig. 10 Correlation between the population of Ω_b and several descriptors for the ethane C_2H_6 molecule as calculated by the DFA set, HF, MP2 and CCSD(T). **A)** Bond length (Å) **B)** Bond Dissociation Energy (kcal/mol) **C)** Harmonic frequencies (cm^{-1}) **D)** Ω_b volumes ($\text{\AA}^3$). Pure functionals are colored blue, hybrid and double hybrid functionals are colored green, and wavefunction methods are colored red. Least-squares fits to the DFA set data are drawn as dashed black lines with their Pearson correlation coefficients shown in each plot.

point. Considering the blue crosses in Figure 11A, corresponding to BX-LYP, it can be clearly seen how a HFX contribution of 10-20% renders the correct population in Ω_b . For this same DFA, good atomization energies are obtained with approximately 20% HFX, while optimal bond lengths (Figure 11B) are obtained at approximately 10% HFX. Therefore, in the case of BX-LYP, atomization energies and bond populations are both optimized at at 10-20% HFX, not far from the composition of the original B3LYP DFA and other general-purpose DFAs. Higher percentages lead to a total breakdown. As per the strong collinearity of bond lengths and harmonic frequencies, the same behavior applies. In the case of SX-VWN3, accurate atomization energies require much higher HFX content, up to 40%, and are not in line with the optimization of the bond population. This exemplifies the problems for describing inhomogeneous densities that plague local DFAs: the correct energetic description can be artificially obtained by tuning the adiabatic connection, but a proper bonding density is not achievable within this parametric space.

In spite of these observations, many successful DFAs with high HFX content have been proposed for a number of good reasons, achieving in most cases remarkable accuracy for thermodynamics and energetics. Smoothing the effect of higher exact exchange contributions by fitting to reference energies may potentially solve this issue for ΔE_{atom} , perhaps at the cost of other features, including the electron density of the bonding region. Indeed, atomization energies are relative and, as such, affected starkly by the description of core electrons of high multiplicity isolated atoms. From the energetic point of view, self-interaction

error is maximal for core regions,³⁹ and therefore including a higher HFX fraction can in principle improve results specially on those systems. Be it through parametrization or due to a physically sound reason, atomization energies are unavoidably prone to error cancellations. It must be noted that reaction energy barriers, another common test for DFAs, depend on the correct description of transition structures, which may be particularly delocalization (self-interaction) error sensitive as well.⁴⁰

In any case, as it has been shown, atomization energies are ill-suited to discuss densities, as they are not strictly speaking an equilibrium property, but rather a property that takes both the molecule and the isolated atoms into account. Better test sets ought to be feasible if relative energies are calculated between systems where the error falls within the same range. Future work will be devoted in this direction.

2.3 The role of bonding densities

While higher amounts of HFX lead to higher populations in Ω_b , this generally leads to lower atomization energies. However, both distances and harmonic frequencies point to stronger bonds. In the previous section some plausible explanations have been suggested, but the effect of increasing the bonding density can be investigated explicitly. To do so, a series of Density Corrected DFT (DC-DFT) calculations have been performed on the test systems. The use of inconsistent densities has been thoroughly justified both from a conceptual and a practical point of view recently.⁴¹

Self-consistent BLYP densities have been used in combination with the BX-LYP family to assess the effect of HFX in the DFA

Fig. 11 Correlation between the population of Ω_b (in electrons) of the N_2 molecule and its **A)** atomization energy (ΔE_{atom}) in kcal/mol, **B)** equilibrium bond length in Å. BX-LYP, in blue, consists on a BLYP modification with increasing amounts of HFX (10-90%). SX-VWN3, in green, is a SVWN3 modification with increasing amounts of HFX (10-90%). Arrows point in the direction of increasing HFX. Least-squares fits to each series are drawn as dashed colored lines. Pearson correlation coefficients (R) are -1.00 for every fit.

when the bonding population is not affected starkly. The resulting methods have been named BLYP//BX-LYP. On the other hand, self-consistent densities from the one-parameter hybrid BX-LYP DFA (for which self-consistent results are shown in Figure 11) have been coupled to the BLYP and B3LYP DFAs. As it has been previously shown, these densities have increasingly populated bonding basins, scaling linearly with their HFX percentage. The resulting methods have been termed BX-LYP//BLYP and BX-LYP//B3LYP respectively. This type of procedure allows for the calculation of energetics and other descriptors while non-selfconsistently increasing the bonding population, thus isolating the effect of the density from the DFA. Geometries have been conveniently optimized in all cases using tight numerical gradients.

Dissociation curves for the N_2 molecule, calculated with different methods, are shown in both Figure 12 and Figure 13. Similar results for the CO molecule can be found in the ESI† (Figures S6 and S7, Table S3). Figure 12 showcases the aforementioned effects of increased HFX DFAs: atomization energies shift significantly, but also importantly, equilibrium distances shrink and potential well amplitudes (in this case, exemplified by fitted Morse potentials, details of which can be found in the ESI†) decrease significantly. In fact, the highest HFX method deviates noticeably from the Morse potential fit, having a sharper descent into the well, unlike the low-HFX or pure analogues. This further underscores the effect of HFX in the bonding picture: bonds that are not only shorter but also more resilient to elongation. As expected, in all cases functional errors are predominant and the use of the BLYP density barely perturbs the curves with respect to the BX-LYP self-consistent densities.

Figure 13 shows, in a different scale, the effect of different bonding densities with the same DFA. As previously suggested by Figure 12, the N_2 system seems to be what has been termed

Fig. 12 Dissociation curves for the N_2 molecule calculated with different methods. Data points have been fitted to a Morse potential. BLYP and B3LYP are shown in black solid and dashed lines respectively, and BX-LYP variants (B10%LYP, B50%LYP and B90%LYP) are shown in red, blue and green dashed-dotted lines respectively. ΔE_{rel} is the relative energy in atomic units with respect to the isolated atoms.

Fig. 13 Dissociation curves for the N_2 molecule calculated with different methods. Data points have been fitted to a Morse potential. BLYP and B3LYP are shown in solid and dashed lines respectively. Methods evaluated on self-consistent densities are shown in black, while BX-LYP densities (B10%LYP, B50%LYP and B90%LYP) are shown in red, blue and green respectively. Inset highlights the nearly perfect overlap between BLYP and B3LYP and their 10% variants. ΔE_{rel} is the relative energy in atomic units with respect to the isolated atoms.

as normal (i.e., it does not suffer from a significant density-driven error), and thus the description given by a DFA does not change significantly upon slightly varying $\rho(r)$. However, it is also evident that the effect of increasing the bonding density is akin to increasing the HFX contribution, shortening bonds and reducing the well amplitude. It must be noted that the GGA BLYP exhibits significantly greater sensibility to the input density than the hybrid B3LYP. In this sense, BLYP can be considered more abnormal for this specific system. On the other hand, the B3LYP DFA is hardly different when evaluated on its own self-consistent density, the B10%LYP one, or in the B50%LYP one. In both BLYP and B3LYP the difference between 10% and 50% HFX is negligible (even more so in B3LYP), yet the difference between 50% and 90% is noticeable.

This is particularly interesting considering that the bonding population in Ω_b given by these densities scales linearly with the HFX content and with the atomization energy ΔE_{atom} , as seen in Figure 11. HFX content and Ω_b density may be linearly related, but the bonding population has a quadratic effect on the amplitude of the Morse well. Therefore, to correct density-driven errors a qualitatively different density should be used, such as the HF one, as hinted in previous works.⁴²

Importantly, and as already highlighted in other works, density driven errors may be superior to those related strictly to the energy evaluation of the DFA for some abnormal systems. As shown in Figures 12 and 13, even in normal systems where the density-driven error is small, the relative depth of a given valley on a potential energy curve can change by tuning the density. The effect of density-based corrections may be even higher if vibrational energies are considered, as in the calculation of free energy profiles with anharmonicity corrections, and specially so in the cal-

ulation of precise kinetic constants where recrossing effects are considered implicitly or explicitly.⁴³

3 Conclusions

Different chemical regions (e.g. bonds, cores, lone pairs) of the electron density are affected by different errors, which may be indirectly improved by parametrization towards energies or geometries. However, this has been shown to lead to incorrect densities. Global descriptors of density errors, such as Λ_{diff} (Equation 4), are inherently biased towards core electrons that are relatively inert. As such, they are not correlated with other absolute measurements of the quality of the density, as the dipole moment. A chemically meaningful perspective, as the one outlined in this work using real space descriptors, is desirable.

Density errors concentrated in covalent bonding regions are important in relative terms. Integrated covalent bonding densities increase with HFX content in a given DFA form, which exerts an attractive electronic force towards the involved atoms. This may wrongly lead to steeper wells in the potential energy surface and shorter bonds, pointing to stronger *intrinsic* bonding energies. In heteroatomic polyatomic systems, this artificial strengthening may be compensated by other unsuspected deformations. An in-depth examination of the advantages and disadvantages of global hybrid DFAs with high HFX contributions, still widely in use, may be due. Indeed, different properties lead to different optimal HFX, and thus different densities and chemistries. Subsequently, caution is advised in the development and usage of flexible range-separated functional forms.

Finally, the usage of inconsistent densities has been shown to be plausible for correcting errors in DFT calculations. The use of the HF density, for instance, has been shown to tighten and strengthen covalent bonds, which may help correct otherwise wrong frequencies. The authors hope that a general understanding of what density to use with a given DFA to correct a given error can be achieved.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

RL gratefully acknowledges Dr. Francesca Peccati for fruitful discussions on practical and philosophical matters, and ED 388 for a PhD grant. VP and RL express their appreciation for the financial support of MINECO/FEDER project CTQ2015-67366-P.

Notes and references

- 1 J. P. Perdew and K. Schmidt, *AIP Conference Proceedings*, 2001, **577**, 1–20.
- 2 J. Sun, A. Ruzsinszky and J. P. Perdew, *Phys. Rev. Lett.*, 2015, **115**, 036402.
- 3 M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew and K. A. Lyssenko, *Science*, 2017, **355**, 49–52.
- 4 P. Verma and D. G. Truhlar, *Physical Chemistry Chemical Physics*, 2017, **19**, 12898–12912.
- 5 D. S. Ranasinghe, A. Perera and R. J. Bartlett, *The Journal of Chemical Physics*, 2017, **147**, 204103.

- 6 P. D. Mezei, G. I. Csonka and M. Kállay, *Journal of Chemical Theory and Computation*, 2017, **13**, 4753–4764.
- 7 K. R. Brorsen, Y. Yang, M. V. Pak and S. Hammes-Schiffer, *The Journal of Physical Chemistry Letters*, 2017, **8**, 2076–2081.
- 8 K. Martin, *Angewandte Chemie International Edition*, 2017, **56**, 5396–5398.
- 9 J. A. Dobado, H. Martínez-García, Molina and M. R. Sundberg, *Journal of the American Chemical Society*, 1998, **120**, 8461–8471.
- 10 M.-C. Kim, E. Sim and K. Burke, *Phys. Rev. Lett.*, 2013, **111**, 073003.
- 11 A. K. Theophilou, *The Journal of Chemical Physics*, 2018, **149**, 074104.
- 12 A. Nagy, *The Journal of Chemical Physics*, 2018, **149**, 204112.
- 13 V. Polo, E. Kraka and D. Cremer, *Molecular Physics*, 2002, **100**, 1771–1790.
- 14 Q. Zhao and H. J. Kulik, *Journal of Chemical Theory and Computation*, 2018, **14**, 670–683.
- 15 K. P. Kepp, *Science*, 2017, **356**, 496.2–496.
- 16 T. Gould, *Journal of Chemical Theory and Computation*, 2017, **13**, 2373–2377.
- 17 P. L. A. Popelier and F. M. Aicken, *CHEMPHYSCHEM*, 2003, **4**, 824–829.
- 18 H. Schmider and A. Becke, *Journal of Molecular Structure: THEOCHEM*, 2000, **527**, 51 – 61.
- 19 H. L. Schmider and A. D. Becke, *The Journal of Chemical Physics*, 2002, **116**, 3184–3193.
- 20 A. D. Becke and K. E. Edgecombe, *The Journal of Chemical Physics*, 1990, **92**, 5397–5403.
- 21 B. Silvi and A. Savin, *Nature*, 1994, **371**, 683.
- 22 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 23 E. R. Johnson, S. Keinan, P. Mori-Saánchez, J. Contreras-García, A. J. Cohen and W. Yang, *Journal of the American Chemical Society*, 2010, **132**, 6498–6506.
- 24 W. Yang and J. Contreras-García, *Acta Physico-Chimica Sinica*, 2018, **34**, 567.
- 25 D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, 1995, **103**, year.
- 26 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, year.
- 27 R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford and C. D. Sherrill, *Journal of Chemical Theory and Computation*, 2017, **13**, 3185–3197.
- 28 L. Tian and C. Feiwu, *Journal of Computational Chemistry*, 2011, **33**, 580–592.
- 29 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Å. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision D.01*, Gaussian Inc. Wallingford CT 2009.
- 30 T. J. Venzani, *Journal of Chemical Education*, 1981, **58**, 423.
- 31 P. Mori-Sánchez, A. J. Cohen and W. Yang, *Phys. Rev. Lett.*, 2008, **100**, 146401.
- 32 M. Rahm and K. O. Christe, *ChemPhysChem*, 2013, **14**, 3714–3725.
- 33 F. Fuster, A. Sevin and B. Silvi, *The Journal of Physical Chemistry A*, 2000, **104**, 852–858.
- 34 B. Silvi, *The Journal of Physical Chemistry A*, 2003, **107**, 3081–3085.
- 35 S. Berski, A. J. Gordon and Z. Latajka, *The Journal of Physical Chemistry A*, 2014, **118**, 4147–4156.
- 36 J. Munárriz, R. Laplaza and V. Polo, *Molecular Physics*, 2019, **117**, 1315–1324.
- 37 J. S. Muentner, *Journal of Molecular Spectroscopy*, 1975, **55**, 490–491.
- 38 V. Polo, E. Kraka and D. Cremer, *Theoretical Chemistry Accounts*, 2002, **107**, 291–303.
- 39 A. Nakata, T. Tsuneda and K. Hirao, *The Journal of Physical Chemistry A*, 2010, **114**, 8521–8528.
- 40 J. L. Bao, L. Gagliardi and D. G. Truhlar, *The Journal of Physical Chemistry Letters*, 2018, **9**, 2353–2358.
- 41 A. Wasserman, J. Nafziger, K. Jiang, M.-C. Kim, E. Sim and K. Burke, *Annual Review of Physical Chemistry*, 2017, **68**, 555–581.
- 42 Y. Kim, S. Song, E. Sim and K. Burke, *The Journal of Physical Chemistry Letters*, 2019, **10**, 295–301.
- 43 H. O. Pritchard, *The Journal of Physical Chemistry A*, 2005, **109**, 1400–1404.