A Bond Charge Model Ansatz for Intrinsic Bond Energies: Application to C–C Bonds

Rubén Laplaza, Victor Polo, Julia Contreras-García

HAL Id: hal-02985750
https://hal.sorbonne-universite.fr/hal-02985750v1

Submitted on 2 Nov 2020

A Bond Charge Model Ansatz for Intrinsic Bond Energies: Application to C-C Bonds.

Rubén Laplaza,[*,†,‡] Victor Polo,[†,‡] and Julia Contreras-García[†]

†*Sorbonne Université, CNRS, Laboratoire de Chimie Théorique, LCT, F 75005 Paris, France*

‡*Departamento de Química Física, Universidad de Zaragoza, 50009 Zaragoza, Spain*

E-mail: rlaplaza@lct.jussieu.fr

Abstract

A simple Bond Charge Model is proposed to predict *intrinsic* bond energies. Model parameters can be derived from the topology of the Electron Localization Function and optimized geometries through classic considerations. Results for carbon-carbon covalent bonds are shown to be very accurate in different chemical environments. Insight can be extracted from the application of the model due to its elementary construction and simple mathematical formulation. The remarkable robustness of the fitted model highlights how different Density Functional Approximations relate geometries, densities and energies.

Introduction

The chemical bond is usually regarded as the central concept in chemistry.[1] However, its utility stems from fuzziness, and defining or probing chemical bonds is far from a solved problem. Analogously, the chemical energy associated with forming or breaking a chemical bond is awkwardly hard to define in a general fashion.

1

The IUPAC states[2] that the bond energy (BE) is "the average value of the gas-phase bond dissociation energies (usually at a temperature of 298 K) for all bonds of the same type within the same chemical species." While this definition is suitable from an experimental and thermodynamical point of view, it clearly misses some desirable properties, as it equates the strength of bonds that should be very different on an individual level. The $C-H$ BE energy in methane is, according to this definition, $\frac{1}{4}$ of the atomization energy of methane. However, this does not necessarily mean that each bond breaking step leading to $C+4H$ is associated with this BE: a more refined perspective is desirable.

Furthermore, in quantum mechanics it is non-trivial to consider a local entity, such as a chemical bond, and isolate it from the total wave function of the system. Accordingly, Bond Dissociation Energies (BDEs) are far more used in benchmarking and testing efforts within quantum chemistry. BDEs can be computed as the energy difference between two points in the free energy surface. However, the BDE includes both the *intrinsic* BE, an equilibrium property related to the strength of the bond, and a certain reorganization energy, RE, that accounts both for geometric and electronic relaxation that takes place after or while the bond breaks. Quite often in chemistry, RE governs over BE in the determination of the thermodynamic pathways: whenever a bond is broken, fragments reorganize to achieve significant stabilization.

Occasionally, separating RE and BE is reasonably feasible. For instance, and ignoring finite basis set issues for the moment, the electronic BE of ethene can be calculated as the electronic energy difference between $H_2C=CH_2$ in its closed-shell singlet state and the two H_2C fragments, with the same geometry, in their triplet $^3B_1$ state (Scheme 1 A). The triplet multiplicity is formally suitable for the formation of the double bond in ethene. In a symmetric covalent system, polarization effects can be assumed to be small. In this case, RE is small compared to BE, as the geometry and electronic state for the methylene fragments are close to the actual minima. However, the electronic BE of ethyne must in principle be computed with respect to the $^4\Sigma^-$ quartet state of the methylidine radical, which differs in
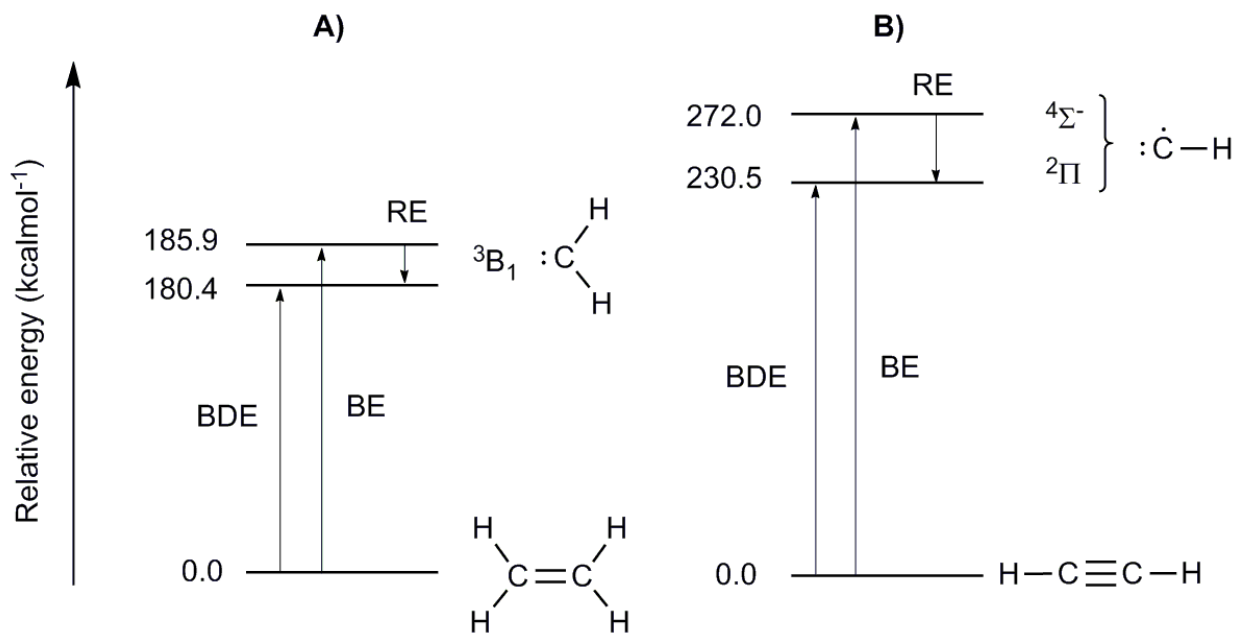
Scheme 1: Energy levels for the homolitic bond breaking of A) ethene into 2 CH_2 fragments, with and without geometrical relaxation in the $^3B_1$ state and B) ethyne into 2 CH fragments in the $^4\Sigma^-$ state and relaxed to the $^2\Pi$ state.

energy from the $^2\Pi$ doublet minima (Scheme 1 B) by more than 40 kcal/mol.

Intrinsic BEs, which we shall refer to as simply BEs, are in principle an equilibrium property; something hidden on the molecular wave function that characterizes the fuzzy entity we call a chemical bond. Nevertheless, both chemical bonds and BEs are invaluably important for the rationalization of chemical phenomena.[3–6] Indeed, relating BEs and BDEs to geometric or electronic features is widespread in the chemical community,[7,8] in spite of a number of troublesome conceptual issues present in different approaches.[9]

Strictly speaking, equilibrium properties, no matter how obscure, should definitely be predictable on the basis of other equilibrium properties. In this work, the Bond Charge Model (BCM) ansatz, based on the electron density, the kinetic energy density and geometrical considerations will be presented. This approach allows for the calculation of BEs for C$-$C bonds within quantum-mechanical accuracy. A posteriori theoretical considerations and considerable insight can be extracted from the application of the model, examples of which will be given in the latter sections.

Computational methods

A set of 59 different Density Functional Approximations (DFAs) in Kohn-Sham Functional Theory, Hartree-Fock (HF) and Frozen Core Möller-Plesset second order perturbation theory (MP2) methods have been used as defined in the Gaussian 09 rev.D01 package.[10] The total array 61 methods will be referred to as the DFA set with numbers *1* to *61*; the detailed list can be found in the Supporting Information (SI). A quadruple-ζ def2-QZVP basis set has been used for all calculations.[11] All (*intrinsic*) BE calculations have been performed preserving the geometry of the fragments as in the original molecule, and adjusting electronic states to match the proper bond order using the unrestricted formalism. No counterpoise corrections have been used as per the large size of the basis set. Spin contamination and basis set superposition errors were monitored and validation calculations can be found in the SI. Solid state calculations of diamond were performed with VASP 5.4.1[12–15] using the HSE06 functional[16] with a 550 eV plane wave cutoff. Integration of electron density within ELF basins was performed with the `CRITIC2` program[17,18] using the algorithm proposed by Yu and Trinkle.[19] Statistical analysis was performed with the Statsmodels package version 0.9.0.[20] Graphics were generated with Chimera 1.11.2[21] and VESTA 3.4.1.[22]

Approximate Energy Models of Covalent Bonds

The Bond Charge Model ansatz

The original Bond Charge Model (BCM)[23–26] is a simple proposal by Parr and coauthors in which the potential energy curve W for an homonuclear diatomic molecule with internuclear distance R is modeled as a sum of three terms (Equation 1): D, which is a bond-specific constant accounting for the internal energy of the participating atoms, V, which is a Coulomb term, and T which is a kinetic energy term. Chemical bonds are assumed to be well described by a negative electronic charge or bond charge, q, between two nuclei with positive charge Z.

Electrostatic forces lead to a $1/R$ dependency for V (Equation 2), and a q/R^2 dependency for T which arises from the particle in a box model (Equation 3). Thus, W can be expressed in terms of a coefficient of the nuclei-bond potential C_V^{BN}, a coefficient of the nuclei-nuclei potential C_V^{NN}, and a bond kinetic energy coefficient C_T^B.

$$W = D + V + T = D + C_V^{BN}\frac{q}{R} + C_V^{NN}\frac{1}{R} + C_T^B\frac{q}{R^2} \tag{1}$$

$$V = \frac{(Z^2 - 4qZ)}{R} = \frac{Z^2}{R} - 2\frac{qZ}{R/2} = C_V^{NN}\frac{1}{R} + C_V^{BN}\frac{q}{R} \tag{2}$$

$$T = \frac{qh^2}{8m_e\nu^2 R^2} = C_T^B\frac{q}{R^2} \tag{3}$$

Where h is Planck constant, m_e is the mass of the electron, and ν is a variable such as $\nu R = R_{bond}$, the effective bond length. Assuming that q must compensate nuclear charges Z to the net charge, the only model parameters are q and ν.

The BCM analytical expression of W allows for the advantageous calculation of equilibrium bond lengths R_{eq}, harmonic oscillator strengths k_e, and bond energies BE (Equations 4-6) among others:

$$(\partial W/\partial R)_{R=R_{eq}} = C_V^{BN}q + C_V^{NN} + 2C_T^B\frac{q}{R_{eq}} = 0 \tag{4}$$

$$(\partial^2 W/\partial R^2)_{R=R_{eq}} = \frac{2}{R_{eq}^3}(C_V^{BN}q + C_V^{NN} + 3C_T^B q/R_{eq}) = k_e \tag{5}$$

$$W_{R=R_{eq}} - D = W_{R=R_{eq}} - W_{R=\infty} = -BE = C_V^{BN}\frac{q}{R_{eq}} + C_V^{NN}\frac{1}{R_{eq}} + C_T^B\frac{q}{R_{eq}^2} \tag{6}$$

While the expressions in Equations 1, 2 and 3 can be used to fit experimental R_{eq} and BE values,[27] without additional constraint the model has led to unphysical parameters which hamper interpretation.[26] Note that several local minima could coexist in the parametric space of the original model. On the other hand, were q known, a perfected fit could be performed using the rightmost expressions in terms of C_V^{BN}, C_V^{NN} and C_T^B. Fitting is justified in the

case of C_T^B in order to capture the unknown value of ν. In principle, C_V^{BN} and C_V^{NN} are related by $C_V^{NN} = (C_V^{BN})^2/16$. However, an independent scaling factor for each term can be added, although the weight of the bond-nuclei term is expected to be higher than the nuclei-nuclei term both because its an electrostatic interaction over a shorter distance and because, after all, bonding does take place. This additional degree of freedom can be understood as a shielding effect: charges are not strictly punctual, particularly so in bonds. Therefore, the effective Coulomb term for the C_V^{BN} and C_V^{NN} terms should be different.

Defining and calculating the bond charge q is not trivial. It has been suggested that the localized depiction of chemical bonding given by this model resembles the partitioning of real space given by the Electron Localization Function (ELF).[28,29] In fact, including the extension to heteroatomic bonds which is not needed here, all model parameters can be defined and obtained from the topology of the ELF: The bond charge, q, associated with a covalent bond A$-$B, is obtained by the integration of the electron density $\rho(r)$ over the space delimited by the disynaptic ELF basin $\Omega_{A,B}$, and nuclear charges Z correspond to the integral over the core basins of the bonded atoms, Ω_A and Ω_B respectively, plus the nuclear charge. The effective bond path, νR_{eq}, can be considered the length of the bond basin $\Omega_{A,B}$. Consequently, it can be calculated by subtracting the radii of the two core basins from the total internuclear distance. Multiple bonds may lead to multiple disynaptic basins, each accounting for an electron pair.

The methodological aspects of the conjoined ELF-BCM have been extensively covered elsewhere,[30] and shown to be valid for covalent bonds. It must be noted that the charge equalization between Z and q may not be sufficient in the context of the ELF-BCM, as in general core populations are constant with respect to the chemical environment but bond population q may migrate towards lone pairs or other bonds in polyatomic systems which are not directly involved in the bonding interaction.

Kinetic energy models

Some limitations of the model can be seen by starting from a simple Morse potential, which is known to be accurate for most dissociation curves. Rewritten, so as to explicitly present an attractive and repulsive term (Equation 7), it becomes apparent that the equilibrium well results from the difference in the slopes of two exponential terms. The positive term, which dominates at small R, has a slope $\propto 2D_e e^{\beta(R_{eq}-R)}$, while the negative term has a slope $\propto -2D_e e^{-2\beta(R_{eq}-R)}$. These two slopes are equal in value when $R = R_{eq}$: when $R < R_{eq}$ the negative term dominates and at larger $R > R_{eq}$ the positive one does. In the BCM ansatz, in principle stabilization arises from the Coulombic $C_V^{BN}\frac{q}{R}$ term, while destabilization comes from a combined $C_V^{NN}\frac{1}{R} + C_T^B\frac{q}{R^2}$ term, including kinetic and electrostatic contributions (*cf.* Eqs.1-3). The ratio of the slopes of the repulsive and attractive terms in the BCM is thus proportional to $(C_V^{NN}R + 2qC_T^B)/qC_V^{BN}R$. The first term $C_V^{NN}/qC_V^{BN} < 1$ for the attractive term to eventually dominate: the bonded state is stable. However, for most of R and specially in constrained bonds, the leading term is $2qC_T^B/C_V^{BN}R$, which highlights the critical importance of the kinetic term in the correct description of the bonded state.

$$W = D_e(1 - e^{-\beta(R-R_{eq})})^2 = D_e(1 + e^{-2\beta(R-R_{eq})} - 2e^{-\beta(R-R_{eq})}) \tag{7}$$

As it is, the expression for T in the original BCM is not particularly justified. In the ELF-BCM framework, another expression for T, now represented by the kinetic energy of a bonding basin Ω_B, has been proposed[30] based on the kinetic energy density of an homogeneous electron gas (Equation 8). The original $1/R^2$ dependence is unaltered.

$$T = -\frac{4\pi C_F}{3\nu^2}\frac{q^{5/3}}{R^2} \propto C_T^{\Omega_B}\frac{q^{5/3}}{R^2} \tag{8}$$

where C_F is the Fermi constant. This model kinetic energy still suffers from some of the deficiencies of the original term. Integrating a more sophisticated inhomogeneous kinetic energy density within Ω_B should lead to better results, yet our goal is simplicity. Gradient

expansions of the kinetic energy density of the homogeneous electron gas involve square gradients of the density. The integral of such quantities over a volume is related to the surface integral of the electron density gradient over the interbasin area. Therefore, two novel kinetic energy terms have been devised that contain a parameter-free damping of q that scales with R while preserving the limits of the original model at $R \to \infty$ and $R \to 0$; stemming from the assumption that, as the distance increases the interbasin surface of the bond must increase or decrease following an unknown dependency. The first one, in Equation 12, is meant to dissipate the charge as the bond expands; the proposal in Equation 13 attains the opposite, dissipating the charge when the bond is contracted. A priori, which behavior is more representative starting from our crude approximation is not known. Similar damping expressions can be constructed using the Gauss error function and its complementary as far as $R > 0$ in all the parametric space.

Therefore, five different expressions for BE have been constructed, out of which Equations 9 and 10 use the original kinetic energy term without and with the shielding improvement, Equation 11 implements the kinetic energy as in Equation 8, and Equations 12 and 13 implement two different opposite damping terms.

$$BE_1 = C_V^{BN} \frac{q}{R_{eq}} + (C_V^{BN})^2 \frac{1}{16 R_{eq}} + C_T^{\Omega_B} \frac{q}{R_{eq}^2} \tag{9}$$

$$BE_2 = C_V^{BN} \frac{q}{R_{eq}} + C_V^{NN} \frac{1}{R_{eq}} + C_T^{\Omega_B} \frac{q}{R_{eq}^2} \tag{10}$$

$$BE_3 = C_V^{BN} \frac{q}{R_{eq}} + C_V^{NN} \frac{1}{R_{eq}} + C_T^{\Omega_B} \frac{q^{5/3}}{R_{eq}^2} \tag{11}$$

$$BE_4 = C_V^{BN} \frac{q}{R_{eq}} + C_V^{NN} \frac{1}{R_{eq}} + C_T^{\Omega_B} \frac{q^{\frac{5}{3}acot(R_{eq})}}{R_{eq}^2} \tag{12}$$

$$BE_5 = C_V^{BN} \frac{q}{R_{eq}} + C_V^{NN} \frac{1}{R_{eq}} + C_T^{\Omega_B} \frac{q^{\frac{5}{3}atan(R_{eq})}}{R_{eq}^2} \tag{13}$$

In the case of symmetric C$-$C bonds, both $\int_{\Omega_C} \rho(r)dr$ and the size of Ω_C remain constant,

as the core of the carbon atoms does not participate. Therefore, both Z and ν are constants, and so the effective bond path νR_{eq} does not represent bond strength or bond order, which in turn ought to be captured by q.

Results and discussion

Application to carbon-carbon bonds

In order to properly parametrize the model, a wide selection of C−C bonds has been considered, covering different bond orders, relative strengths, geometrical constrains and keeping possible polarization or non-covalent effects to a minimum. The following molecules have been considered for this purpose: C_3H_6 (**1**), H_3C-CH_3 (**2**), C_6H_6 (**3**), $H_3C_4-C_4H_3$ (**4**), $H_2C=CH_2$ (**5**), $HC\equiv C-C\equiv CH$ (**6**), $HC\equiv CH$ (**7**) and C_2 (**8**). A set of fluorinated analogues have been included for validation purposes, considering the interest some of these molecules arise.[31–33] It is the case for the *gauche-* and *trans-* conformations of 1,2-Difluoroethane (**9** and **10**), the *cis-* and *trans-* isomers of 1,2-Difluoroethylene (**11** and **12**), and Tetrafluoroethylene (**13**). All calculated molecules are shown in Figure 1. The composition of the set is meant to minimize reorganization phenomena and non-bonded interactions and hence provide fairly accurate BEs. BEs, internuclear equilibrium distances and Ω_{CC} populations were extracted accordingly.

For the model to work, it follows that bond order and geometrical strain should be taken into account by the combination of q and R_{eq}. As shown in Figure 2 A and B for simple molecules, both parameters are in fact known to be weak descriptors of bond strength that have been previously used in the context of C−C bonds.[34] Collinearity with each other is not high (Figure 2 C).
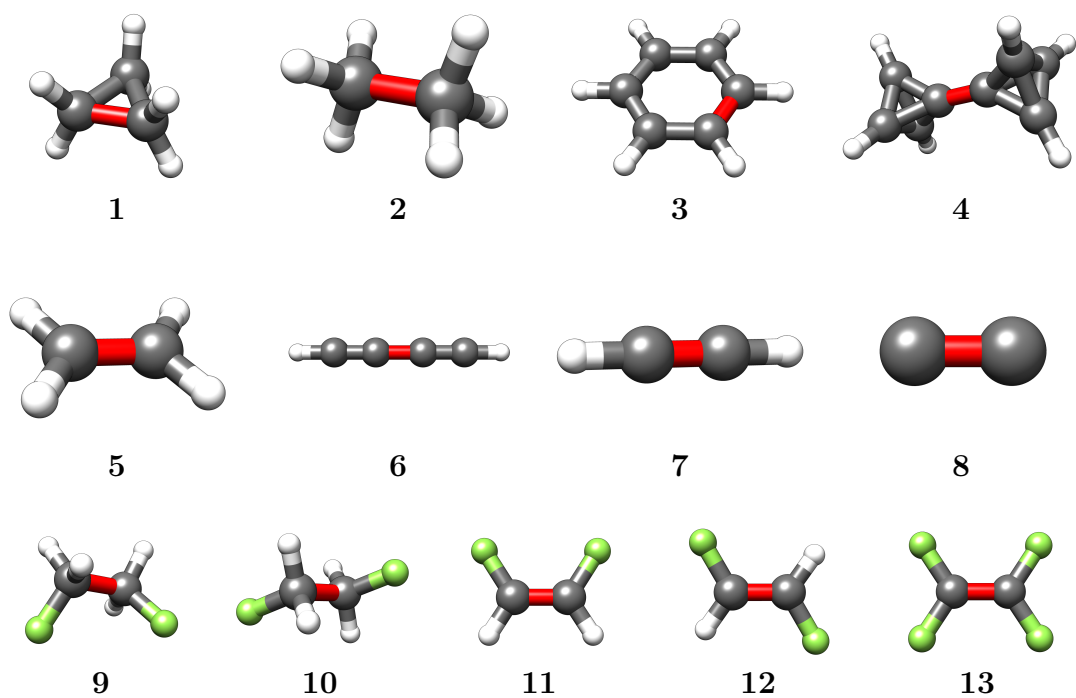
Figure 1: Calculated molecules in the parametrization set. Hydrogen, carbon and fluorine atoms are depicted as white, grey and light green balls-and-sticks respectively. Relevant C−C bonds are colored red.
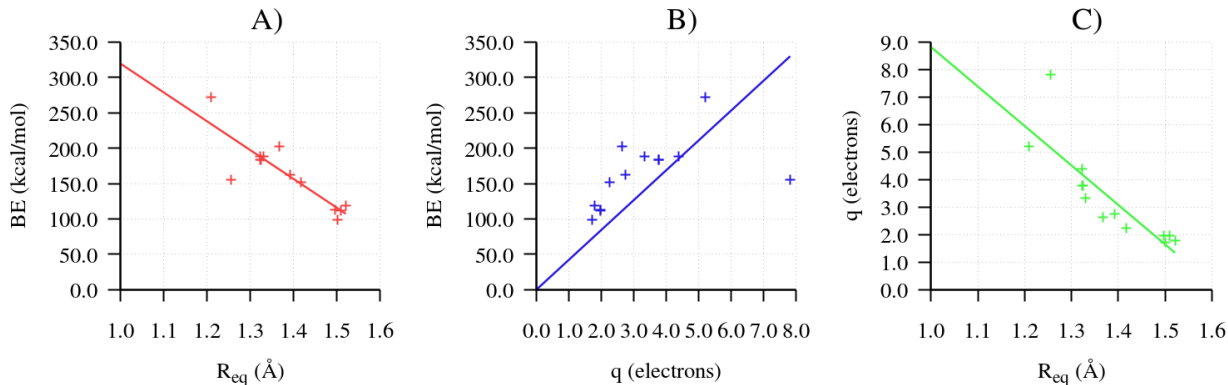
Figure 2: Linear correlations between R_{eq}, q and BE for the C−C bonds in the molecule set. All calculations performed at the MP2-FC/def2-QZVP level.

As a first approach, the coefficients were obtained by using the results of the ωB97XD DFA, which is widely considered appropriate when dealing with covalent bonds in organic systems.[35] Correlation coefficients and parameters for the different models, BE_1 to BE_5 (see Equations 10 to 13) are collected in Table 1. Without imposing any additional restraints, only models 3 and 4 seem to accommodate the physical meaning we want, in which C_V^{BN} is the only positive (stabilizing) contribution to BE. They also present the best regression coefficients and $F - statistic$ results, while the kurtosis result for BE_4 is significantly closer to a normal distribution. The inclusion of the electrostatic shielding degree of freedom from model 1 to model 2 is shown, on the basis of the adjusted R^2 value, to be meaningful. Similarly, the kinetic energy term derived for model 3 achieves a noticeable improvement over the initial one, although it does increase the kurtosis significantly. As predicted and hinted in previous work,[30] such term is apt for non-strained bonds, hence outliers are found among the test-set. In particular, Cook's distance for the C_2 molecule in model 3 was found to be 12.861 while no other molecule has a value superior to 0.1. Thus, the inclusion of the damping term at short distances in model 4 improves the value of R^2 to a quantitative level by smoothing the anomalous C_2 molecule, while preserving the correct physical interpretation and limits. Similar conclusions can be drawn from other reduced data sets, as reported in the SI.

Table 1: Multivariate ordinary least-squares regression parameters C_V^{BN}, C_V^{NN} and $C_T^{\Omega_B}$ for models BE_1 to BE_5 (see Equations 10 to 13). The $F-statistic$ is calculated with respect to the trivial $C_V^{BN} = C_V^{NN} = C_T^{\Omega_B} = 1$ model.

Model	C_V^{BN}	C_V^{NN}	$C_T^{\Omega_B}$	$F-statistic$	$Kurtosis$	R^2	$Adjusted R^2$
BE_1	-51.1359	2614.8789	-52.6352	80.0	3.260	0.936	0.924
BE_2	-236.1163	305.9522	278.2517	72.2	2.352	0.956	0.943
BE_3	150.9831	-42.5211	-39.0838	146.7	6.626	0.978	0.971
BE_4	247.6594	-120.5810	-96.6753	401.1	3.519	0.992	0.989
BE_5	-26.6361	232.8136	25.2538	54.9	2.203	0.943	0.926

The root mean square (RMS) error of BE_4 is 4.2029 kcal/mol, which is in the same order of magnitude that it can assume for a given DFA. Concluding that BE_4 is our best approximation, a robust linear model was used to fit the same data. The fitting was performed by iterative re-weighted least-squares using Huber weights. Relatively close parameters $C_V^{BN} = 229.7770$, $C_V^{NN} = -96.2613$ and $C_T^{\Omega_B} = -90.2421$ were obtained, all within the standard errors of the coefficients of the ordinary least-squares model, which corroborates that outliers or high-leverage molecules are not substantially over-weighted. Note that the conditions in Equations 4 and 5 are not enforced here with respect to the reference data because the underlying physics (i.e. preserving the dependency of the terms) and the applicability (reproducing the calculated BEs) are deemed more important. For instance, for a given equilibrium pair of R_{eq} and q to be the BE maxima of model 3, the following equality must hold (Equation 14):

$$C_T^{\Omega_B} = \frac{R_{eq}(C_V^{NN} - C_V^{BN}q)}{2q^{5/3}} \tag{14}$$

while the case for the best fit, model 4, is somewhat harder (Equation 15):

$$C_T^{\Omega_B} = \frac{3R_{eq}(1 + R_{eq}^2)(C_V^{NN}q^{-\frac{5}{3}atan(R_{eq})} - C_V^{BN}q^{1-\frac{5}{3}atan(R_{eq})})}{6 + 6R_{eq}^2 - 5R_{eq}\ln q} \tag{15}$$

Consequently, the derived parameters produce biased dissociation curves with respect to R. High fidelity dissociation curves could be obtained by enforcing the expressions above

with a fixed bond charge q. Such possibilities may be explored in future work.

General trends of the model

Model 4 (Equation 10) was used to fit an extensive database (793 datapoints) including the complete set of DFAs using an ordinary least-squares model. Results are shown in Figure 3a and Figure 3b with respect to the model terms and q and R_{eq} respectively. The collective model takes $C_V^{BN} = 232.2327$, $C_V^{NN} = -107.4466$ and $C_T^{\Omega_B} = -88.9169$, reasonably close to previous results, with $R^2 = 0.985$. Various methods introduce additional variability, thus the kurtosis value increases to 7.036. The RMS error is slightly increased to 5.7890 kcal/mol. Details and diagnostics of the fitting procedure can be found in the SI. Overall, the model fits the data at a semi-quantitative level, which is remarkable considering its simplicity and the variability in the data set. However, it must be noted that the ensemble parameters reported above are not recommended for in-depth analysis: method-specific parameters should be used instead, which lead to $R^2 > 0.99$ in most cases, and can be found in the SI for a selection DFAs.
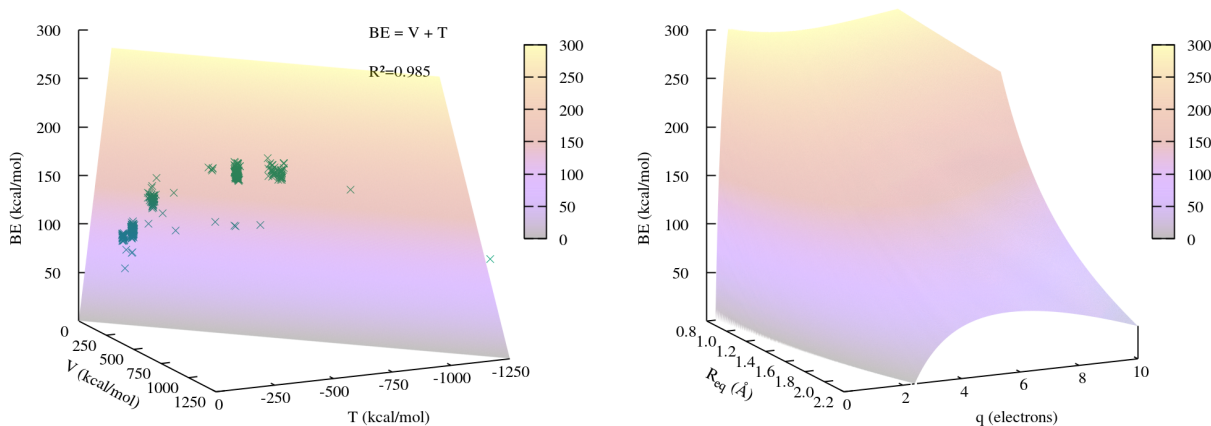
(a) BE with respect to T and V   (b) BE with respect to q and R_{eq}

Figure 3: Predicted BEs with respect to (a) T and $V = V_{BN} + V_{NN}$ and (b) q and R_{eq}. Reference data points in (a) are shown in green. All energies in kcal/mol.

The high quality of the obtained fit reveals major similarities in the way different DFAs relate geometries, ELF-given Ω_B populations (q) and BEs. In average, all molecules suit the model irrespective of the calculation method. Average absolute values of standardized residuals per molecule and method are presented in Figures 4 and 5 respectively. Values per molecule range from 0.2084 for cyclopropane to 2.1310 for ethyne (**7**). No particular issues arise for any of the strained systems; the fact that ethyne is the upper limit is reasonable considering that it is the only triple bond in the parameter set and still the standardized residuals are not extremely worrying. C2 (**8**) has a large standard deviation in this average, pointing at the difficulties of capturing its correlated character in DFT: different DFAs give largely different depictions. Method-wise, values range from 0.5113 for Xα to 2.7600 for HF, which is to be expected. The kinetic term is a correction on top of an homogeneous electron gas derivation, and so simple DFAs should provide good fits, even if the predicted BEs are not accurate, while HF is not based on the electron density and suffers from very inconsistent correlation treatment for different molecules. Most DFAs are evenly treated, as
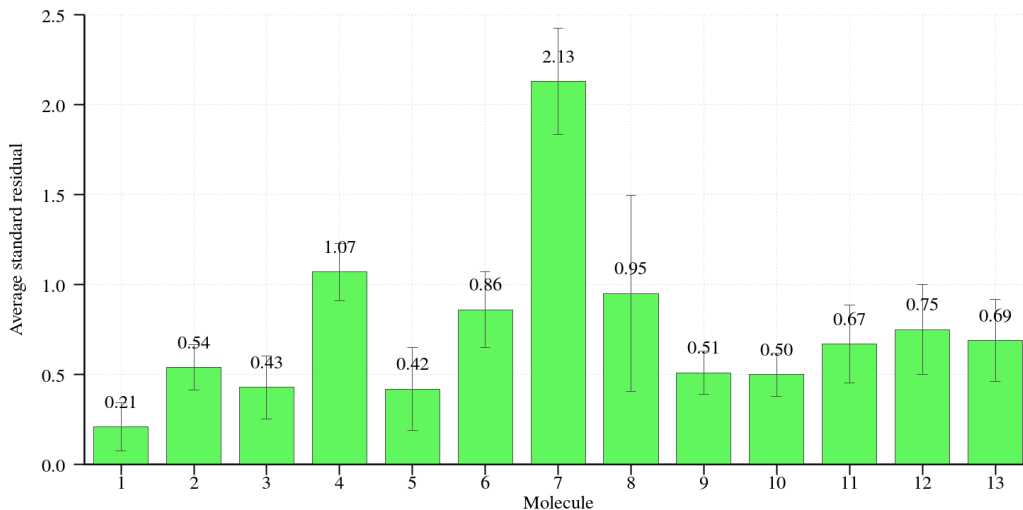
14

shown in Figure 5.

Figure 4: Averaged absolute value of standardized residuals per molecule in the test set. Molecule **7** corresponds to ethyne, molecule **8** corresponds to C_2. Error bars spawn one standard deviation for the given averages.
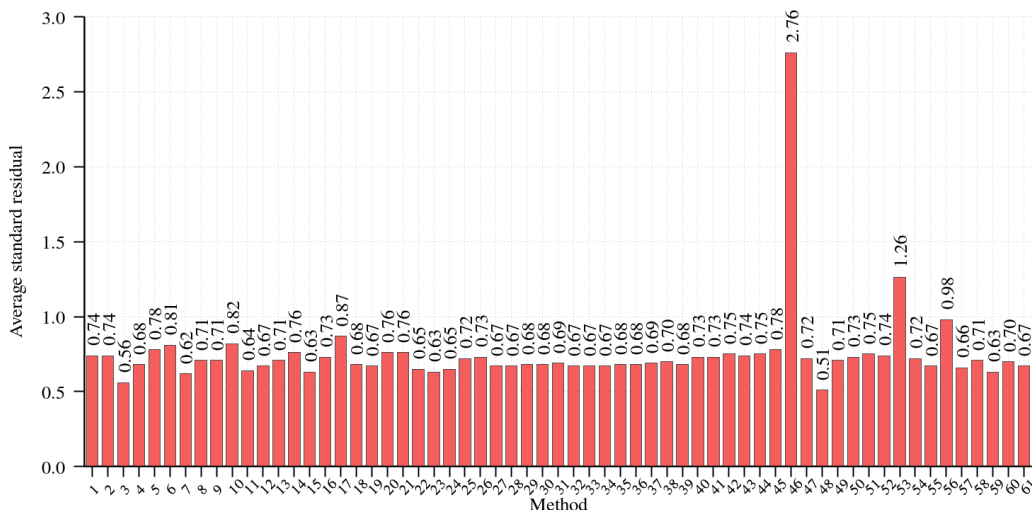

Figure 5: Averaged absolute value of standardized residuals per method in the test set. Method *46* corresponds to HF, method *48* corresponds to the Xα DFA.

As per the fitted curve, general trends can be ascertained. For a fixed q, the model resembles a dissociation energy curve in an approximate way, as seen in Figure 6. Note that for any given q, predicted energy minima correspond with extremely short R. In reality, q changes with R, therefore the energy quickly tends to 0 at long R as the bond charge

dissipates. Stronger bonds with higher values of q are seen to be sensitive to R. Interestingly, there is an estimation for the maximum BE for a given q.
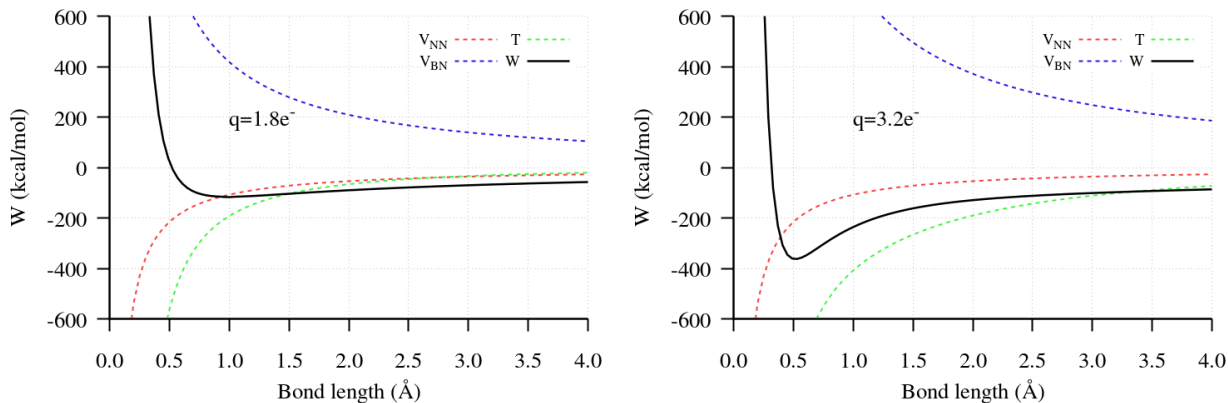

Figure 6: Evolution of V_{NN}, V_{BN} and T with R according to the fitted model parameters (Model 4) for a fixed value of q.

At any given fixed R_{eq}, there is a number of shared electrons q that gives a maximal BE, as shown in Figure 7, which becomes steeper as bond length increases. Note that all the fluorinated species are perfectly reproduced in the model even when isomers have extremely similar R_{eq} values. This signals that the electronic effects are perceived in the C−C bond through q and reflected in BE. Note that while C_2 remains somewhat of an outlier, it is reasonably well fitted by the model. This approves the usage of the whole valence density of the molecule as bond charge q, which corresponds to a formal quadruple bond. It is important to note that the somewhat outlier nature of ethyne and dicarbon is related to their particular nature but also influenced by the lack of other molecules with formal bond orders between 3 and 4 in the test set. This makes the statistical model inherently biased towards the lower bond orders, even through robust regression techniques.
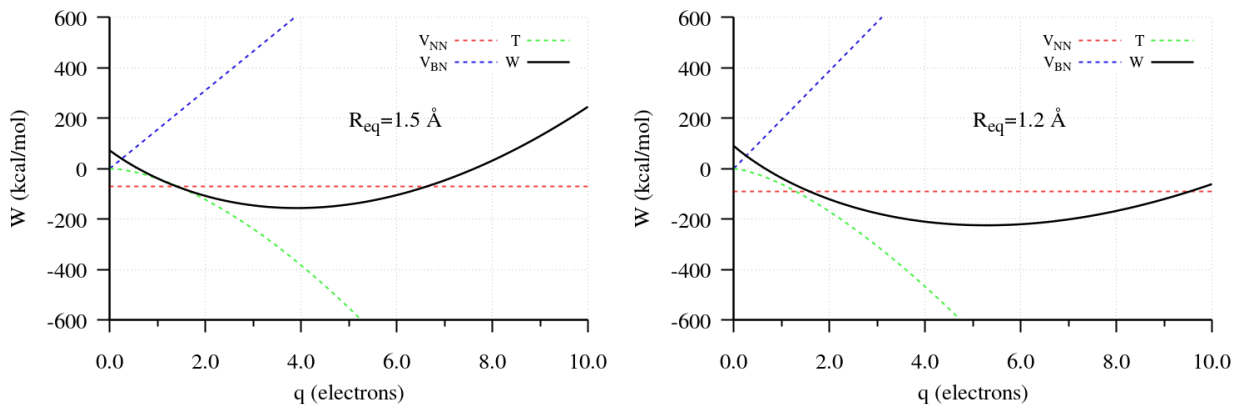
Figure 7: Evolution of V_{NN}, V_{BN} and T with q according to the fitted model parameters (Model 4) for a fixed value of R_{eq}.

Example applications

Using DFA-specific parameters, the proposed model is able to recover BEs from equilibrium properties R_{eq} and q. While extrapolating results towards high q and low R (formal bond orders >3) may not be accurate, interpolation within the single-to-double bond regime should provide highly accurate results. This is given by the current parametrization set, which is focused on single and double bonds, and the subsequent fact that residuals are largest for C_2 and ethyne. A simple application is comparing the BE of $C-C$ bonds of the cyclopentadienyl anion (Figure 8a) per se and in ferrocene $Fe(C_5H_5)_2$ (Figure 8b) in order to evaluate the differences in strength of the carbon structure upon coordination.
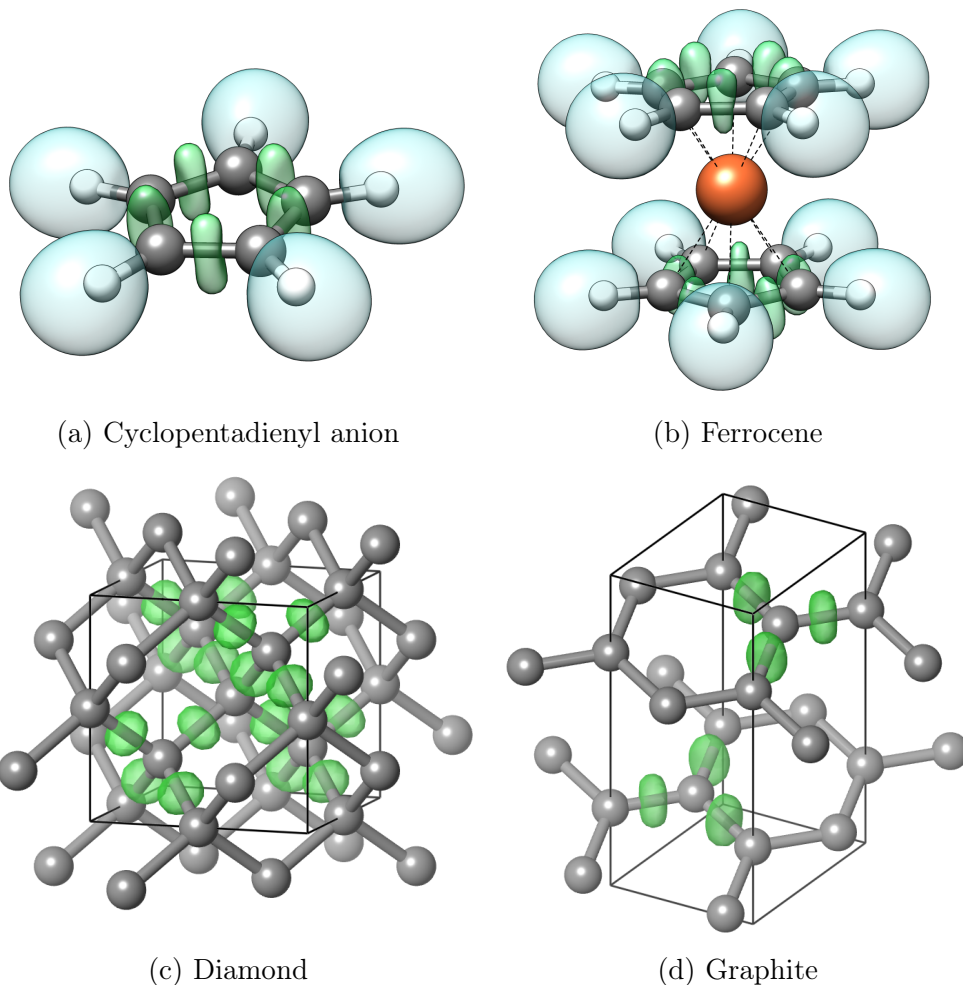
(a) Cyclopentadienyl anion

(b) Ferrocene

(c) Diamond

(d) Graphite

Figure 8: ELF depictions of (a) the cyclopentadienyl anion , (b) ferrocene $Fe(C_5H_5)_2$, (c) diamond and (d) graphite with an isovalue of 0.85. Carbon, hydrogen and iron atoms shown as grey, white and orange balls-and-sticks respectively. $C-C$ basins are depicted in green, while $C-H$ basins are colored light blue.

The optimized geometries and wave functions for both systems at the ωB97XD/def2-QZVP level give $q = 2.9002$ and $R_{eq,C-C} = 1.4067$ Å for the anion and $q = 2.5414$ electrons and $R_{eq,C-C} = 1.4182$ Å for ferrocene. Coherently, a large amount of charge is transferred to the formally doubly charged iron atom. In spite of this, BEs of 159.88 and 146.12 kcal/mol respectively are obtained, which means that each $C-C$ bond becomes 13.76 kcal/mol weaker in ferrocene. Note that the BE of $C-C$ bonds in benzene is 159.62 kcal/mol at this theory level, which is almost the same result for the anion. In fact, bond-nuclei attraction V_{BN} is increased in the anion with respect to the neutral benzene, as given by the extra electron, but

most of the stabilization is counteracted by an increase in T, rendering the same BE. When comparing to ferrocene, the small loss in BE per bond (mostly in form of V_{BN}) is globally justified by the formation of the haptic bond with the metallic center, and explains why the carbon structure remains similarly stable in the coordinated species. Hence, no significant improvement in C−C cleavage is in principle expected. Note that the difference in strength is in agreement with the changes in the stretching C−C frequency from ferrocene, around 1400 cm^{-1}, compared to benzene, around 1500 cm^{-1}.[36] Similar analysis may be performed on transition structures and molecular systems of interest in the context of computational organic and inorganic chemistry.

Analogously, the presented model may be used to study solid state systems, among which carbon allotropes are highly important. Optimizing the geometry of diamond using the HSE06 DFA (Figure 8c) provides a C−C equilibrium distance $R_{eq,\text{C−C}} = 1.5359$ Å and the integration of the valence electron density in the ELF basins gives localized covalent bonds with $q = 2.0052$. With the HSE06 parameters (available in the SI) a BE of 113.02 kcal/mol is obtained, which is analogous to the simple bond in ethane in the set (112.93 kcal/mol). Note that this result suggests that covalent bonds in diamond are very similar to simple C−C bonds elsewhere, with any energetic differences arising purely from reorganization terms. Interestingly, the natural decomposition of the proposed ansatz shows that the alleviation of nuclear repulsion V_{NN} in diamond is able to counteract the loss of bond-nuclei stabilization. Graphite, on the other hand, has $R_{eq,\text{C−C}} = 1.4154$ Å and $q = 2.6667$ electrons, which estimates a BE of 148.42 kcal/mol, in the order of weaker aromatic bonds. This supports the notion that aromaticity in graphite is slightly different from smaller aromatic hydrocarbons, which has been suggested previously.[37]

Conclusions

In this work a Bond Charge Model is developed based on ELF topological analysis and a revised kinetic energy term. Fitted to reference data, the model predicts the BE of $C-C$ bonds from equilibrium properties, achieving quantitataive results within DFT accuracy.

The advantages of the ELF-BCM ansatz, its extreme simplicity and straightforward interpretation, are therefore available for the analysis and interpretation of quantum mechanical results at a similar level of uncertainty. Further developments in this line should enable a complete toolkit connecting topological descriptors with energetic features. For instance, coupling with $C-H$ potentials should provide access to atomization energies of hydrocarbons. Modelling potential energy curves or anharmonic frequencies is also possible through this model. The foundations presented in this work may also be of interest for constructing more sophisticate machine learning models, yet machine learning may not always be the answer to regression problems.

Some examples of applications in organometallic and solid state chemistry are provided in order to illustrate the ease of application and the availability of insight, which we expect may be of interest in computational studies. The variables of the model may also be obtained experimentally.

Acknowledgement

Supporting Information Available

Regression data and coefficients for different DFAs. Calculation details and a complete list of methods.

References

(1) Ball, P. *Nature* **2011**, *469*, 26–28.

(2) *IUPAC Compendium of Chemical Terminology*; IUPAC.

(3) Grimme, S. *Journal of the American Chemical Society* **1996**, *118*, 1529–1534.

(4) Howard, S. T.; Cyranski, M. K.; Stolarczyk, L. Z. *Chem. Commun.* **2001**, 197–198.

(5) Exner, K.; Schleyer, t. . T. j. . T. v. . . n. . . p. . . y. . . d. . j. U. . h. e. . h., Paul von Ragué

(6) Brenner, D. W.; Shenderova, O. A.; Harrison, J. A.; Stuart, S. J.; Ni, B.; Sinnott, S. B. *Journal of Physics: Condensed Matter* **2002**, *14*, 783–802.

(7) Zavitsas, A. A. *The Journal of Physical Chemistry* **1987**, *91*, 5573–5577.

(8) Martín Pendás, A.; Francisco, E. *Phys. Chem. Chem. Phys.* **2018**, *20*, 16231–16237.

(9) Kaupp, M.; Danovich, D.; Shaik, S. *Coordination Chemistry Reviews* **2017**, *344*, 355 – 362, Chemical Bonding: State of the Art.

(10) Frisch, M. J. et al. Gaussian 09, Revision D.01. Gaussian Inc. Wallingford CT 2009.

(11) Weigend, F.; Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *119*.

(12) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47*, 558–561.

(13) Kresse, G.; Furthmüller, J. *Computational Materials Science* **1996**, *6*, 15 – 50.

(14) Kresse, G.; Joubert, D. *Phys. Rev. B* **1999**, *59*, 1758–1775.

(15) Kresse, G.; Hafner, J. *Journal of Physics: Condensed Matter* **1994**, *6*, 8245–8257.

(16) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. *The Journal of Chemical Physics* **2006**, *124*, 219906.

(17) de-la Roza, A. O.; Blanco, M.; PendÃąs, A. M.; LuaÃśa, V. *Computer Physics Communications* **2009**, *180*, 157 – 166.

(18) de-la Roza, A. O.; Johnson, E. R.; LuaÃśa, V. *Computer Physics Communications* **2014**, *185*, 1007 – 1018.

(19) Yu, M.; Trinkle, D. R. *The Journal of Chemical Physics* **2011**, *134*, 064111.

(20) Seabold, S.; Perktold, J. Statsmodels: Econometric and statistical modeling with python. 9th Python in Science Conference. 2010.

(21) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *Journal of Computational Chemistry* **2004**, *25*, 1605–1612.

(22) Momma, K.; Izumi, F. *Journal of Applied Crystallography* **2008**, *41*, 653–658.

(23) Parr, R. G.; Borkman, R. F. *The Journal of Chemical Physics* **1967**, *46*, 3683–3685.

(24) Borkman, R. F.; Parr, R. G. *The Journal of Chemical Physics* **1968**, *48*, 1116–1126.

(25) Parr, R. G.; Borkman, R. F. *The Journal of Chemical Physics* **1968**, *49*, 1055–1058.

(26) Borkman, R. F.; Simons, G.; Parr, R. G. *The Journal of Chemical Physics* **1969**, *50*, 58–65.

(27) Pasternak, A. *Chemical Physics* **1977**, *26*, 101 – 112.

(28) Becke, A. D.; Edgecombe, K. E. *The Journal of Chemical Physics* **1990**, *92*, 5397–5403.

(29) Silvi, B.; Savin, A. *Nature* **1994**, *371*, 683.

(30) Munárriz, J.; Laplaza, R.; Martín Pendás, A.; Contreras-García, J. *Phys. Chem. Chem. Phys.* **2019**, *21*, 4215–4223.

(31) Goodman, L.; Gu, H.; Pophristic, V. *The Journal of Physical Chemistry A* **2005**, *109*, 1223–1229, PMID: 16833433.

(32) Banerjee, D.; Ghosh, A.; Chattopadhyay, S.; Ghosh, P.; Chaudhuri, R. K. *Molecular Physics* **2014**, *112*, 3206–3224.

(33) Andrada, D. M.; Casals-Sainz, J. L.; MartÃŋnâĂ ĚPendÃąs, Ã.; Frenking, G. *Chemistry âĂŞ A European Journal* **2018**, *24*, 9083–9089.

(34) Chesnut, D. *Chemical Physics* **2001**, *271*, 9 – 16.

(35) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(36) Latouche, C.; Palazzetti, F.; Skouteris, D.; Barone, V. *Journal of Chemical Theory and Computation* **2014**, *10*, 4565–4573.

(37) Popov, I. A.; Bozhenko, K. V.; Boldyrev, A. I. *Nano Research* **2012**, *5*, 117–123.

Graphical TOC Entry