Cite this: DOI: 00.0000/xxxxxxxxxx

New venues in electron density analysis

Bruno Landeros-Rivera,[*a] Miguel Gallegos,[b] Julen Munárriz,[b] Rubén Laplaza,[a] and Julia Contreras-García[*a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

We provide a comprehensive overview of the chemical information within the electron density: how to extract information, but also how to obtain and how to assess the quality of the electron density itself. After introducing several indexes derived from the electron density which allow to reveal bonding, we focus on the various potential sources of the electron density, and also explain on the error trends they show so that a judicious choice of methods and limitations are clearly laid on the table. Computational, experimental-computational combinations, as well as machine learning efforts, are covered in this spirit.

1 Introduction

Chemical sciences stem from an elegant interpolation scheme rooted in the periodic table. Properties and reactivity of matter can often be rationalized and even predicted from chemical composition and structure. For instance, it is reasonable to guess, solely based on the orbital energies of the H and C atoms, that hydrocarbons will in general not be very polar and require a large energy input to react. To improve upon this first, crude guess, chemists require information about the chemical structure, linked to composition through our understanding of chemical bonding and valence: once we know how the atoms in such hypothetical hydrocarbon are arranged, one can assign bond orders and bond types, identifying relevant motifs (e.g. aromatic rings, π- interactions, etc..) that lead to an even better predictive power.

From the point of view of quantum mechanics, this is only natural given that composition plus structure govern the external potential of the system, $v(\mathbf{r})$. As the coordinates and nuclear charges fully determine – in the non-relativistic Born-Oppenheimer framework – the ground state many-body wave function of the system, a significant part of chemistry can be conceived as the phenomenological study of the changes of the external potential. The remaining information needed to fully characterize a chemical system is the number of electrons, N, with which the exact Hamiltonian and the exact ground state energy E can be obtained. Both changes in $v(\mathbf{r})$ and N are usually merged into chemical modifications. It is at this level that most empirical phenomena are understood, ranging from acid–base behavior to solubility, to name a few. For instance, the effect of substituting H by F in an organic molecule often causes predictable changes in its properties, which can be explained in terms of relative (atomic) electronegativities and the resulting bond ionicity. Ultimately, these can be traced back to the delicate balance between the changes in nuclear charge and electrons, which are coarsely captured by the trends of the periodic table.

As the energy is the key property of a chemical system, leading both to time evolution and to correspondence to macroscopic behavior, let us re-frame the aforementioned analysis mathematically starting from the derivative of the ground state energy E, which can be written as

$$dE = \left(\frac{\partial E}{\partial N}\right)_{v(\mathbf{r})} dN + \int \left(\frac{\delta E}{\delta v(\mathbf{r})}\right)_{N} \delta v(\mathbf{r}) d\mathbf{r} \tag{1}$$

Whereas the first term of right hand side of Equation 1 represents the chemical potential μ, the second one provides the electron density $\rho(\mathbf{r})$:

$$\rho(\mathbf{r}) = \left(\frac{\partial E}{\partial v(\mathbf{r})}\right)_{N} \tag{2}$$

Hence, understanding how these two quantities change, is crucial for the study of chemical structure and reactivity. Notwithstanding the complexity of the chemical potential, the 3D nature of the electron density leads to convoluted relationships, difficult to analyze and summarize.

In what follows, we will discuss the information content of the electron density and its relation with other density-derived quantities that can be obtained both experimentally and computationally. We will then introduce different approaches for obtaining electron densities, trying to go beyond the mere computational approach. The relevance of the electron density clashes with the little attention its accuracy has received in the computational world until very recently. Thus, a section will be devoted to recovering the different types of electron densities that can be used

[a] *Laboratoire de Chimie Théorique, Sorbonne Université, F. 75005 Paris, France; contrera@lct.jussieu.fr*
[b] *Departamento de Química Física y Analítica, Universidad de Oviedo, 33006 Oviedo, Spain.*

in the analysis and their intrinsic errors. We will end with a little summary of the perspectives within electron density analysis that these different studies open.

2 Chemical information from the electron density

In a first approximation, the electron density of a system can be seen as a sum of exponentially decaying functions centred at the nuclei, one exponential per shell for each of the composing atoms. This approximation, known as promolecular, is also used as a first approximation to fit structure factors in crystallographic studies and as an initial guess in *ab initio* electronic structure computations alike.

When atoms bond together, the exponentially decaying tails of their corresponding atomic electron density cusps intersect. According to Coulomb's law, an accumulation of electron density in the internuclear region should ease the repulsion between the two nuclei and be strongly stabilizing;[1–4] at the same time, the exact many-body electron density is correlated in terms of spin and electron-electron interactions.[5]

In general, it can be said that the deviation from the initial promolecular density is directly related with interactions between atoms.[6] As such, it has become customary to study the electron density via the analysis of its critical points, as popularized in the Quantum Theory of Atoms in Molecules (QTAIM) introduced and developed by Bader and coworkers.[7–10] For practical reasons, we shall assume that the analysis is performed on a continuous and differentiable field homeomorphic to $\rho(\mathbf{r})$ where the nuclear cusps are replaced by sharp maxima. In the QTAIM framework, all critical points where $\nabla\rho(\mathbf{r}) = \partial\rho/\partial x + \partial\rho/\partial y + \partial\rho/\partial z = 0$ are first located and then characterized by computing the corresponding second derivatives at the coordinates of the critical point $\mathbf{r}$, as in

$$H(\rho,\mathbf{r}) = \begin{bmatrix} \partial^2\rho/\partial x^2 & \partial^2\rho/\partial x\partial y & \partial^2\rho/\partial x\partial z \\ \partial^2\rho/\partial y\partial x & \partial^2\rho/\partial y^2 & \partial^2\rho/\partial y\partial z \\ \partial^2\rho/\partial z\partial x & \partial^2\rho/\partial z\partial y & \partial^2\rho/\partial z^2 \end{bmatrix} \tag{3}$$

where $H(\rho,\mathbf{r})$ is the Hessian matrix of ρ at $\mathbf{r}$. Diagonalizing $H(\rho,\mathbf{r})$ yields three eigenvalues λ_1, λ_2 and λ_3. In QTAIM parlance, the standard notation for critical points is (r,g), where the rank r is the number of non-zero eigenvalues λ_i, and the index g is defined as $\sum \lambda_i/|\lambda_i|$ (i.e., the number of negative eigenvalues). Chemical bonds are associated in the literature to (3,-1) points; but there are other relevant points, such as (3,+1) - ring critical points- and (3,+3) - cage critical points. (3,-1) points are saddle points, which are a maximum in two directions of space, and minimum along the other orthogonal (bonding) direction. It is worth mentioning at this point that maxima ((3,-3) critical points) that do not correspond to nuclear cusps can be found in some situations, such as stretched out of equilibrium bonds.[11,12] Such non-nuclear maxima of $\rho(\mathbf{r})$ have been subjected to significant controversy, and can sometimes be associated with solvated electrons[13] or arising as mere computational artifacts.[14–16]

As $\rho(\mathbf{r})$ is continuous and differentiable (with the aforementioned exception of nuclear cusps), there will be maximal value paths connecting critical points. This way, the so-called bond paths connect nuclear and bond critical points. The full set of bond paths leads to a graph structure which conveniently retrieves a Lewis-like picture (e.g., valence rules are satisfied for organic molecules). Similar critical points analysis can be performed for all kinds of systems where determining bonding patterns may not be trivial, including periodic lattices.[17] Exemplary representations can be found in Figure 1. Note that the paths do not strictly coincide with the shortest internuclear lines, i.e. the lengths of such paths differ slightly from the internal coordinates, possibly reflecting the influence of strain (see cubane in Figure 1d).[18,19]

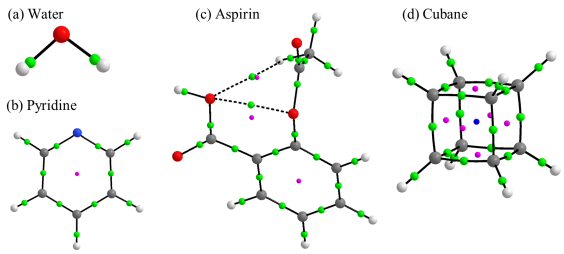

Fig. 1 Critical points of $\rho(\mathbf{r})$ and bond paths that connect them for (a) water, (b) pyridine, (c) aspirin, (d) cubane. CPs are shown as spheres. Bond critical points are colored in green, ring critical points are coloured in purple and cage critical points are colored in blue. Electronic structure calculations were performed at the PBE/def2-SVP level of theory,[20] and topological analysis was carried out by means of the AIMALL package.[21]

Going back to Eq. 1, we can hypothesize that the bundle of critical points and their corresponding electron density values should be a good fingerprint of the exact Hamiltonian of a molecular system. This has been the main focus of QTAIM for a long time: the position and electron density values at the (3,-1) critical points have been typically used to localize and characterize the bond strength of families of bonds.

Going beyond the analysis of critical points, it is also possible to retrieve global descriptors from the analysis of the electron density. The gradient field of $\rho(\mathbf{r})$, $\nabla\rho(\mathbf{r})$, provides the rate of change of the electron density. Surfaces where this gradient becomes null naturally delimit atoms within a molecule. This is a substantial insight of QTAIM[8] that agrees with the previous reasoning. In space, the exponential decay of atomic cusps continue until some interaction is established with some other density distribution; the intersecting surfaces constitute a natural boundary between atoms. Such partitioning also has a number of convenient mathematical properties that we shall not treat here,[22] and is a cornerstone among the many atomic partitioning schemes that have been developed over time.[23]

We may also calculate how fast the density decreases (or increases) w.r.t. a reference density, which is normally taken to be the well-known homogeneous electron gas (HEG) to avoid mismatch due to different local potentials. This allows to define the dimensionless reduced density gradient (RDG, $s(\mathbf{r})$):

$$s(\mathbf{r}) = \frac{1}{C_s}\frac{|\nabla\rho(\mathbf{r})|}{\rho(\mathbf{r})^{4/3}}, \tag{4}$$

where $C_s = 2(3\pi^2)^{1/3}$. By definition, $s(\mathbf{r}) \geq 0$. Low values of $s(\mathbf{r})$

are associated with critical points of $\rho(\mathbf{r})$ but also with very weak intramolecular interactions which do not show an electron density critical point due to structural constrains. Thus, the analysis of $s(\mathbf{r})$ is specially advantageous for studying weak interactions.By inspecting sign$(\lambda_2)^*\rho(\mathbf{r})$, the relative strength and attractive or repulsive nature of interactions can be determined. Note that sign(λ_2) is the sign of the second eigenvalue of the Hessian matrix, and it reflects charge accumulation or depletion in the plane perpendicular to an interaction region. In cases where $\lambda_2 > 0$, the electron density is depleted, which is associated with steric repulsion. On the other hand, $\lambda_2 < 0$ implies that electron density is accumulated, as expected for strong non-covalent interactions such as hydrogen bonds. Finally, regions where $\rho(\mathbf{r}) \approx 0$ are typical of weak dispersive interactions. Isosurfaces of $s(\mathbf{r})$ reveal the presence of non-covalent interactions in 3D space. While strong localized interactions manifest as thick disks between atomic pairs, weak interactions are observed as extended flat surfaces where more than two atoms are involved. In particular, the latter are not properly described by QTAIM because bond critical points are always pairwise. Thus, QTAIM and NCI analysis can complement each other.[24,25]

The Laplacian of the electron density, $\nabla^2\rho(\mathbf{r})$, is the sum of the diagonal elements of $H(\rho,\mathbf{r})$ (3), $\lambda_1 - \lambda_3$ and can also be used to reveal chemical bonding information.[26] The shell structure of atoms and molecules is manifested in $\nabla^2\rho(\mathbf{r})$ as alternating regions of electron charge concentration ($\nabla^2\rho(\mathbf{r}) < 0$) and depletion ($\nabla^2\rho(\mathbf{r}) > 0$), being the outer shell with $\nabla^2\rho(\mathbf{r}) < 0$ named the valence-shell charge concentration (VSCC). When the VSCC of two atoms overlap, it is said that shared interactions exist between them (covalent bonding). The other case corresponds to closed-shell interactions, as in ionic, van der Waals, metallic or hydrogen bond interactions. Figure 2 highlights closed-shell interactions in LiF, where the VSCCs (red isocontours) of both atoms do not overlap, but a region of charge depletion (blue isocontour) is found in the interatomic space. In contrast, shared interactions can be observed in pyridine, where the VSCCs of all covalently bonded atoms overlap. Both types of interactions can be noticed in aspirin, which is stabilized by covalent and non-covalent bonding. The (3,+3) critical points of $\nabla^2\rho(\mathbf{r}) > 0$ found in the VSCC are associated with bonded and non-bonded electron pairs, providing in most of the cases a good correspondence with the VSEPR model.[27]

It should be noted that other density-based indexes have been proposed to describe a wider range of other types of interactions. The Strong Covalent Interaction index (SCI)[28] is able to distinguish between double, triple, quadruple and quintuple covalent bonds by the visual analysis of its isosurface in the center of a bond, which show shapes resembling dumbbells, donuts, four beads, and two corn beans, respectively. Furthermore, the Bonding and Non-covalent Interaction index (BNI) is capable of describing covalent, as well as ionic, metallic and other specific localized interactions such as chalcogen and halogen bonds,[29] by performing a similar visual inspection. Both SCI and BNI are founded in the Pauli energy. Likewise, the Ultra Strong Interaction index (USI), which has a similar expression to NCI (equation 2), with the difference that the norm of the gradient of $\rho(\mathbf{r})$ is replaced by its Laplacian, provides a proper description of systems with bond orders higher than four.[29]

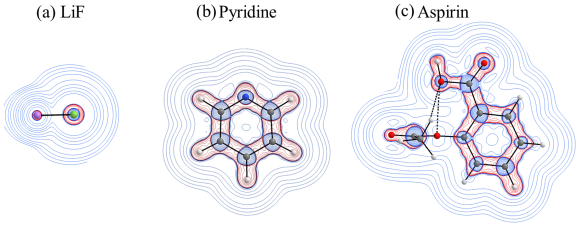

Fig. 2 Isocontours of $\nabla^2\rho(\mathbf{r})$ on planes containing selected bonds for: (a) LiF (ionic system), (b) pyridine (covalent system), and (c) aspirin (covalent system with non-covalent interactions). Negative Laplacian values are depicted in red, while positive ones are shown in blue. Electronic structure calculations were performed at the PBE/def2-SVP level of theory,[20] and topological analysis was made by means of AIMALL package[21]

Finally, we note that quantitative chemical information can also be obtained from $\rho(\mathbf{r})$ in the context of conceptual Density Functional Theory.[30,31] In this discipline, some important chemical concepts associated with reactivity are linked to functions that measure the response of a system to changes in the number of electrons or external potential. For instance, electronegativity is directly related to minus the chemical potential (as defined in Equation 1), while the second derivative of this function with respect to N corresponds to Pearson's hardness.[32] Another important function developed in the framework of Conceptual DFT is the Fukui function,[33] which is useful for detecting the electrophilic and nucleophilic sites of a molecule. We direct readers to Refs.[31,34] for an extended discussion regarding the definition of Chemical Concepts in the framework of conceptual DFT.

3 Other approaches for obtaining the electron density

An adequate representation of the subtle interplay of contributions that dominate chemical interactions often requires accurate molecular densities and expensive computational tools. Regarding the former, the demanding scaling law attributed to the estimation of the electron density has limited the application of quantum and computational chemistry to systems of small to medium size. This is particularly problematic, as many of the very delicate, but crucial, aspects of chemical interactions, such as Non-Covalent Interactions (NCI), can only be properly understood under the magnifying glass of a solid and robust analysis of $\rho(\mathbf{r})$ and its related quantities. Such an inconvenience can be overcome by resorting to other sources of information, such as experiment or machine learning derived electron densities.

3.1 Experiment

One of the main advantages of the electron density is that it is a quantum observable that, unlike the wavefunction, can be obtained experimentally from X-ray diffraction. This is very attractive, since it allows for comparison of experimental and theoretical models. In crystalline systems, the scattering of X-rays is caused mainly by electrons, and it is described by the so-called

structure factors, F_{hkl}. The F_{hkl} are related to the electron density by a Fourier transform, and the $|F_{hkl}|^2$ are proportional to the X-ray scattered intensities, which are the experimentally measured quantities. Unfortunately, since F_{hkl} are in general complex functions, the experiment can only provide direct information about their modulus but not about their phases. Thus, the electron density cannot be determined by simply applying the Fourier transform to the square root of the scattered intensities. As a consequence, the solution of crystallographic structures (which depends on methods that retrieve the lost information) provides only approximate atomic positions that need to be improved by a subsequent process, the so-called crystallographic refinement. The latter consists in generating modeled F_{hkl} that will be refined against the experimental one, normally by using the least-squares method. The number of fitting parameters (refinement parameters) will depend on the chosen model.

Usually, the sum of spherical atomic densities constitutes the first approach in the electron density reconstruction from X-Ray data. This model, equivalent to the aforementioned promolecular density, is known in crystallography as the Independent Atom Model (IAM). IAM has been used in most refinements,[35] surprisingly providing accurate structures in many cases despite its simplicity. Naturally, for a deeper analysis of the properties of a crystal that goes beyond the elementary structural determination, more complicated models based in aspherical electron densities need to be employed. One of the most successful methods for non-spherical refinements is the Hansen-Coppens multipole model (MM).[36] In MM, the electron density is modeled by *deformation functions* which are products of radial functions and spherical harmonics that can be expanded up to fourth order. Besides the atomic coordinates and thermal parameters, the population of these functions, called multipoles, are employed as extra refinement parameters. Further, the valence shell can be expanded or contracted, potentially leading to electron fluctuations that facilitate charge transfer between the atoms. In its more advanced version, the core density can also be expanded/contracted and change its population.[37] As it could be anticipated, MM is superior to IAM in various ways. One of special interest here is that electron densities derivatives are available, so that QTAIM critical points and atomic basins, can be determined within MM.[38]

From a physical point of view, a major pitfall is the non-uniqueness of the models, since different refinements could be performed (with some changes in the multipoles) that will be statistically valid.[39] This could lead to uncertainties in the topological descriptors, in particular in those that are susceptible to subtle changes, such as the Laplacian. It should also be noted that even though MM provides a more faithful representation of the crystalline $\rho(\mathbf{r})$, in the cases where anharmonic motion is substantial the refinement could lead to non-physical multipoles and unrealistic negative electron densities.[40] Moreover, like IAM, MM is not able to provide accurate anisotropic displacement parameters (which describe thermal motion) of H atoms without resorting to additional methods.[41] Also, for these atoms the multipole expansion is normally limited to the first order (dipoles).[42] For these cases, hybrid experimental-theoretical approaches, which will be discussed later, have shown to be advantageous.

3.2 Machine learning

Along the last decades, the use of Machine Learning (ML) in different scientific disciplines has increased dramatically owing to the ability of ML algorithms to perform complex and intricate tasks at a fraction of the standard computational cost. Within this context, the development of Neural Networks (NN), able to approximate any function with arbitrary accuracy[43,44], is reshaping the course of quantum and computational chemistry. Indeed, the blatant success of Deep Learning (DL), and related strategies in the chemistry realm has lead to faster, yet reliable, tools for many different purposes such as property prediction[36,45–49], atomistic simulations[50,51], quantum-chemically accurate force fields[52–54] and potentials[55,56] or novel techniques for chemical discovery and sampling[57–59], to name just a few.

Significant efforts have been devoted in recent years in the computation of accurate densities from molecular geometries[60,61] using ML approaches. Indeed, the fact that ML can be used to map the electron density to the nuclear potential comes to no surprise, as the Kohn-Sham equations ensure the existence of such mapping. However, constructing such models should arguably be more difficult than learning the functional itself, which to a degree explains why more efforts have been devoted to the latter task.[62] There are mainly two reasons for this: firstly, the electron density has a tighter convergence requirement than the energy, meaning that a satisfactory accuracy on the density ought to be more challenging. Secondly, its 3D nature demands an appropriate basis for its expansion over space, for which no universal solution may be available. As a final note, ML efforts targeting the electron density have to specifically tackle the choice of a loss-function,[63] which is difficult owing to the non-trivial comparison of 3D scalar fields. Any ML approach requires a way of quantifying how good or bad the current model is at the task at hand. In the case of $\rho(\mathbf{r})$, this means quantitatively determining how far a predicted density is from the reference density, which is fundamentally undefined because, as it will be discussed later, what makes a good density good is highly debatable (recall that it is a 3D quantity!). This problem is accentuated further due to the approximate nature of the training set, as large databases of exact electron densities are currently unfeasible.

In any case, increasingly accurate ML models of the electron density open the door to a wealth of applications regarding X-ray refinement, DFT, and molecular interpretation.

4 On the spotlight: electron density errors

The electron density is often considered to be robust because it is strongly dominated by nuclear exponential cusps, which in turn guarantee that its general appearance remains constant independently of its origin (quantum chemical computation with different methods, experiment, etc.). However, as one takes derivatives of $\rho(\mathbf{r})$ differences eventually become apparent even between closely related approaches (e.g. upon change of density functional). This is particularly striking given the interpretation of $\nabla\rho(\mathbf{r})$ and $\nabla^2\rho(\mathbf{r})$ explained before, in which such small changes can lead to qualitatively different chemical interpretations. Some authors have pointed out that parametrization strategies in the development of new approximations to the ex-

act exchange-correlation function, by virtue of focusing only in reproducing exact energies, may incur in unphysical errors in $\rho(\mathbf{r})$.[64–68] After all, it is easy to envision a surrogate function that is able to reproduce the behavior of the exact energy in some subspace (notably, integer values of N and Z) while failing to reproduce its derivatives (Eq. 2), hence giving the "right answer" (energy) for the "wrong reason" (density).[69,70] More sophisticated fitting procedures (vide infra) may be able to incorporate a plethora of norms and targets to alleviate these deficiencies. Nevertheless, this brings into question the assumed robustness of $\rho(\mathbf{r})$, which hampers interpretative efforts stemming from it,[71] as well as quantitative uses in DFT, where derivatives of the density play a major role.[72–75]

Indeed, it has been suggested that using "better" densities in order to evaluate the energy non-selfconsistently in the context of DFT offers a way to avoid some known caveats of density functional approximations.[76–78] These approaches are often termed density-corrected density functional theory (DC-DFT), and usually rely on the Hartree-Fock single particle density as their "unbiased" reference. While this strategy has seen some success in challenging systems where standard DFT may fail,[79,80] it immediately leads to questions regarding the overall required quality of the density, in a context in which either experimentally derived or Full Configuration Interaction densities (numerical discretization non-withstanding)[81] provide arguably rigorous yet impractical ground truths. In Layman terms, the question to answer is: how good do we need the electron density to be? And, subsequently: what is the most pragmatic way to obtain such densities? These questions remain open for now.

4.1 How to measure the error

Before analyzing common trends in density errors, it is of undeniable importance to understand how to measure these errors, for instance for the construction of loss functions in ML frameworks (vide supra). Although it is common to use simple scalar metrics, such as the RMSD for the analysis of energy errors, the electron density presents a more difficult situation. Being a 3D field, metrics such as the RMSD (and other statistical aggregates) can hide error compensation.[69] More specifically, the electron density being much higher in the cores, it usually takes the leading term and hides the errors in the valence density, which might be more chemically relevant,[66] and much higher if relative errors are computed (e.g. in the N_2 molecule the intra-method density error near the nuclei is less than 2%, while the error in the valence region is $>10\%$[68]). Without any further consideration, this means that depending on the error metric, a density with a minuscule relative error in the electron cusp of a heavy atom may appear much worse than a density from which all non-covalent interactions (low-density regions) are completely erased, which would in turn poorly describe the interactions in the system. Additionally, many statistical error metrics for $\rho(\mathbf{r})$ can not be compared across different external potentials (i.e. with different methods at their optimized geometries), which introduces an unfair choice of a reference geometry that may mask relevant features.[82] Consequently, by naively using inappropriate error metrics, apparently

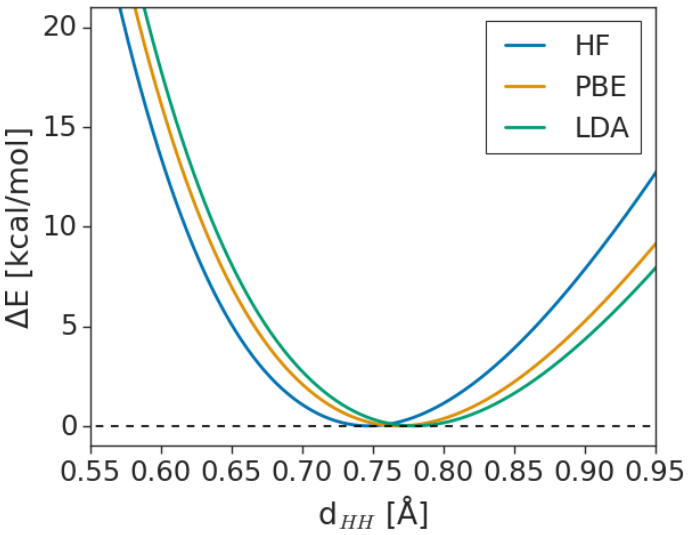

Fig. 3 Dissociation curves of the H_2 molecule computed with three different methods, HF, PBE and LDA, and the def2-SVP basis set.

better performing methods can be inaccurate for predicting properties that depend on the electron density distribution, such as the dipole moment. By resorting to chemically interpretable properties, such as bond charge given by the ELF[83,84] or electron density values at BCPs, it is possible to compare results across different methods at their optimized geometries and focus exclusively in the valence errors.

4.2 Main error trends

It becomes thus of paramount importance to know in advance, what errors we should expect in our working densities before carrying out their interpretation. In general, covalent bonding densities have shown to increase with the exact exchange content, and to lead to steeper wells in the potential energy surface and shorter bonds, both criteria pointing to stronger in situ bonding energies. This phenomena is displayed even in simple cases, such as the bond in the H_2 molecule (Figure 3), but prevails in other covalently bound systems.[68] Indeed, it is a general trend that is directly related to delocalization error:[85] functionals tending to over-delocalize the electron density due to artificial self-interaction between electrons, thus favoring intermolecular space at the detriment of intramolecular covalent bonds. The opposite is observed for HF, which lacks electron correlation contributions. Moreover, it has been observed that local density approximations, plagued by delocalization error, and HF, may represent two extremes in terms of electron delocalization, which in turn may propagate to the calculation of properties (this is the case of the equilibrium bond length of H_2 in Figure 3).[86]

Observing the electron density in bonding regions, the aforementioned trends are once again reproduced: in Figure 4 it can be seen how the density difference between the CCSD and the HF methods in the internuclear region of the H_2 molecule hints at an excessive accumulation of density in the HF result. The opposite trend is observed for LDA, in which the internuclear region is noticeably depleted. From the electrostatic point of view,

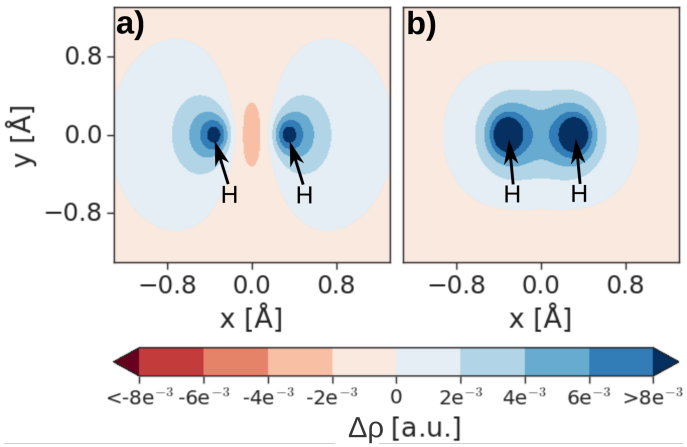

Fig. 4 $\Delta\rho = \rho_{CCSD}$ - ρ_{method} plotted on the X,Y plane of the H_2 for a) *method*= HF and b) *method*=LDA. All calculations performed with the def2-SVP basis set. $\Delta\rho$ has been truncated at $\pm 8e^{-3}$ a.u. to better highlight the valence density.

overly-localized covalent bonds given by inappropriate densities must lead to distorted models of chemical reactivity due to the increased shielding of the electrostatic repulsion between nuclei, which makes such bonds harder to break (e.g. steeper potential curve in Figure 3). However, given the prevalence of DFT in which the connection between energies and densities is approximate, such effects may be masked by error compensation and thus not be reflected in the computed relative energies.

5 New perspectives in electron density information

Two main domains have exploded in the last years, enabling the extraction of more and richer information from the electron density: hybrid theory-experiment approaches that have pushed the limits of crystallographic resolution and machine learning approaches which enable to connect electron density with complex properties, beyond previous models.

5.1 Hybrid experiment-theory approaches

Experimental electron densities have recently seen a great improvement thanks to the introduction of simulations in the refinement (and post-refinement) process. Specially noteworthy is the Hirshfeld Atom Refinement (HAR)[87], where the non-spherical atom form factors are generated by a applying a Hirshfeld partition to an *ab initio* computed electron density, which is obtained by performing a single point calculation at the initial geometry. Then, the atomic coordinates and the thermal parameters are refined with the usual least-squares procedure, and the process is repeated until convergence. One of the most relevant benefits of HAR is its ability to refine anisotropically hydrogen atoms[88,89] which, as mentioned before, is unfeasible with IAM and MM. As an example, the bond critical point corresponding to the intramolecular H···H interaction in biphenyl, and the corresponding ring critcal point (red and yellow spheres, respectively), are compared in Figure 5 for the MM and HAR, as well as for the

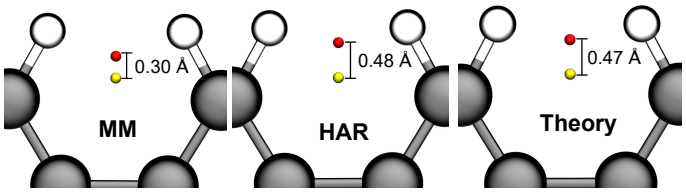

Fig. 5 Bond critical point (red) corresponding to the H···H interaction in biphenyl. The distance to the ring critical point (yellow) is shown in angstroms for MM, HAR and solid-state DFT geometry optimization calculation. Data taken from reference[90]

solid-state DFT optimized geometry of this system (data taken from reference[90]). As it can be seen, MM predicts the structure to be closer to a catastrophe than HAR and the periodical DFT calculation (performed with the same density functional), whose results are essentially the same. As it could be anticipated, this hybrid-theory method is providing outcomes closer to those obtained from conventional quantum mechanical computations.

Following a HAR refinement, it is even possible to obtain an experimental wavefunction using the so called X-ray Constrained Wave function (XCW)[91,92]. In XCW the geometry is fixed and the wavefunction (or Kohn-Sham orbitals) is forced to reproduce the experimental structure factors by minimizing the functional

$$L[\Psi] = E[\Psi] - \lambda(\chi^2[\Psi] - \Delta) \quad (5)$$

where the first term in the right-hand side stands for the energy of the system, and the second accounts for the fitting of the orbitals to the experimental data. The χ^2 is a statistical function that indicates the agreement between the modeled and experimental structure factors, and Δ is the desired level of agreement. The parameter λ, originally considered a Lagrange multiplier, is adjusted during the fitting and gives the strength of the constriction. The application of HAR followed by XCW is know as X-ray Wave function Refinement (XWR).[93,94] By constraining the wavefunction to reproduce the X-ray diffraction data some, deficiencies of using molecular calculations such as the inappropriate description of bulk or correlation effects can be (at least partially) corrected.[95–97] For example, in a recent study XCW was applied to ammonia crystallographic data measured at 160 K, using HF and DFT.[97] It was shown that, while the crystal field tends to localize the electron density in bonding regions, the introduction of correlation through the experimental data does the opposite: it delocalizes the electron density in the inter-molecular regions. Moreover, it was shown that as λ is increased, the HF and DFT electron densities converge (Figure 6), providing further evidence that XCW is able to improve both methods. Thus, these methods offer an opportunity to highlight the errors in the density that are caused by using approximated functionals,[64] and at the same time could offer an alternative approach to correct them. Additionally, consecutive cycles of HAR and XCW can be performed to further improve the model (a process called *total XWR*),[98] and bring the best possible agreement with the experimental data.

Despite these advantages, more research is needed to understand how to minimize the effect of introducing experimental un-

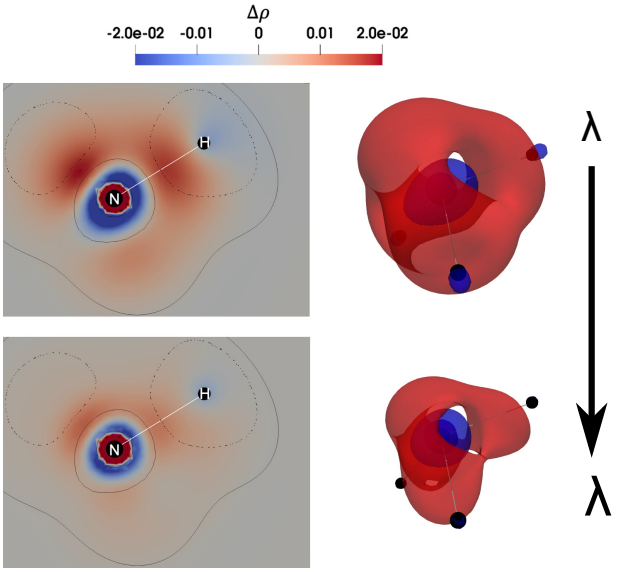

Fig. 6 Plots of the difference in electron density between HF and DFT methods for $XCW_{\lambda=0.001}$ and d) $XCW_{\lambda=0.030}$ for ammonia molecule ($\Delta\rho = \rho(\mathbf{r})_{HF} - \rho(\mathbf{r})_{DFT}$). Isosurfaces with negative values (-0.005 a.u.) in blue and positive values (0.005 a.u.) in red. Figure modified from reference.[97]

certainties ($\sigma(F^2)$) in the computed wave functions and the associated properties.[99] Regarding the latter point, it was shown that the problem of non-uniqueness discussed above for the multipole refinement can also be found in XWR, although for different reasons. The χ^2 function is sensitive to the manipulation of the experimental data, such as the transformation of $\sigma(F^2)$ that is carried out by different crystallographic software. Since the constrained wave function depends directly on χ^2 (equation 5), then the associated physical state will be subjected to the arbitrariness of the method or program employed to process the crystallographic data. For example, for the XWR applied to the crystallographic data of SO_2,[99] it was shown that for a fixed λ value the number and position of the critical points of $\nabla^2\rho(\mathbf{r})$ changed depending of the software used to handle anomalous dispersion, or the method employed to compute $\sigma(F)$ from $\sigma(F^2)$ (the latter is carried out because the measured data and their uncertainties are normally given as F^2_{hkl}). In contrast, the coordinates and thermal parameters obtained from HAR barely change when different refinement strategies are used. Overall, this is currently an active field of research that can bring together the computational and experimental communities, providing the user with more accurate electron densities.

5.2 Electron density descriptors in machine learning

Even though machine learning techniques have been proven to be enormously useful (see Section 3.2), the straightforward application of ML algorithms to chemistry suffers from an important limitation: performing an accurate and representative "chemical featurization". In other words, issues arise from the way in which chemically meaningful information (such as atomic positions, charges, or any specific property) is transformed into a machine-readable format encoding. This is generally done in the form of atomic environment vectors (AEVs), which contain the desired information in a computer-understandable manner. Thus, a lot of effort has been made in recent years within the scientific community to develop suitable featurization approaches[100,101] such as the Bag of Bonds scheme[36], Coulomb matrices[46,102], Atom Centered Symmetry Functions (ACSF)[103,104], along with its many different flavors[105–107], or the more recent Gaussian Moments[108], which would ultimately allow for the construction of reliable ML approaches. Despite rigorous and useful, most strategies still recover little to no information about the actual chemistry, and are usually focused on encoding the radial and angular environments and the chemical composition of the system, just to mimic the external potential without further chemical insight.

The prediction of chemically meaningful quantities may be considerably improved if information of the delicate balance that holds the atoms bounded within a molecule, the chemical bond, is explicitly included. In this context, QTAIM-like descriptors could lead to a more complete feature representation. Although the implementation of the latter is not immediate, given the permutational, translational and rotational symmetries to be fulfilled by standard AEVs, the use of QTAIM properties as ML descriptors could pave the way towards the more accurate prediction of chemical properties. This would lead to potentially interpretable models owing to the direct mapping of such features to the chemical world (as discussed in the previous sections regarding molecular graphs in particular). Indeed, it has been shown that density-based descriptors can be used for the accurate prediction of relative and reaction energies[47,109], proving how very suitable ML chemical features can be distilled within the QTAIM framework. For such purposes, a major caveat is the need for a fully converged electron density, as resulting from experiment or DFT, to construct QTAIM-inspired representations.

For straightforward interpretability, ML techniques have been used within the framework of energetic partitioning schemes, such as the Interacting Quantum Atoms (IQA) approach, leading to the prediction of chemically insightful energetic terms[109–111]. In this context, it is particularly relevant to highlight the work of Popelier and co-workers, who have shown, within the FFLUX project[112], that ML techniques can be very efficiently applied to the atomic partitioning that stems from the QTAIM framework. In this regard, it has been proven[109] that a very accurate (MP2-like level) computation of the dynamic correlation energy is possible using kriging approaches. Analogously, a similar work[113] has been shown capable of optimizing small molecules at a very reasonable computational cost, paving the way towards the development of general truly applicable topological force fields. We envision that further interplay between the topology of $\rho(\mathbf{r})$ and its derived scalar fields and ML methods will see the light in the forthcoming years.

In the more fundamental topic of DFT development research, ML is in a privileged position[114], allowing to help in the design of more physically rigorous, yet practical, DFT functionals; as already stressed in the literature by some authors[115]. The actual development of new functionals such as the DeepDFT[116] and DeepMind 21[117] projects are major examples in which the non-linearity of ML is exploited to construct approximate functionals

with remarkable properties. It is thus clear that ML proposes a crucial tool in the search for new DFT functionals which could, perhaps, conciliate the never-ending dispute between rigor and practicality among DFT developers. However, it is worth noticing that given the interpolating nature of ML approaches, the complex sampling of the chemical space and the black-box nature of some ready-to-use algorithms, special care should be taken to prevent the rise of poorly generalizable and non-interpretable ML-driven functionals.

Finally, in passing it may also be worth mentioning that beyond energetic partitioning schemes, QTAIM in general, can also benefit from the computational boost offered by ML techniques. Indeed, some of us have worked[45] in a NN model capable of predicting the local value of the electron density within a topological atom, in the form of atomic charges. The success of these and similar NN models suggest that the boundaries of QCT in chemistry are likely to vanish in the very near future owing to the remarkable performance of ML-driven quantum chemical tools.

5.3 Density in the Conceptual-DFT framework

Finally, as far as conceptual DFT is regarded, future perspectives are nicely summarized in Ref.[118]. These include an extension of DFT functionals to account for such important effects as temperature, solvent and mechanical forces, as well as time-dependent and excited-state phenomena, or a broader application to kinetic and thermodynamics, among others. Moreover, some functionals derived from information theoretical approaches[119] and whose analytical expression is known, such as the Shannon or the Ghosh–Berkowitz–Parr entropies, have been shown to be promising for the quantitative analysis of relevant properties like regio- and stereo-selectivity, acidity and basicity, aromaticity and conformational stability, to name just a few.

6 Conclusions

Electron density occupies a privileged position in the chemical narrative, holding all the information inherent to any system[120] and its underlying nature. Consequently, a lot of effort has been devoted to its measurement, its computation and its analysis. Nevertheless, these efforts had followed an off-center bias, where the electron density was the mean to something else. On the one hand, computations mainly focused on the electron density as a mean to the energy. Only recently the quantum community has become aware of the relevance of a having a good density, relocating the electron density in the central position it deserves. A similar twist is now being observed in the crystallographic community, which for long was mainly focused on determining correct structures, while paying little attention to the electron density. For many decades it was considered that the inability of refining anisotropically H atoms with X-ray diffraction data was intrinsic to the experiment. Only now that modern techniques like HAR have shown that the problem was (at least in some cases) not a deficit of the reflection data, but a shortcoming of IAM, the importance of using and developing more robust models for the electron density has become significant for the advance of crystallographic methods.

The exploding machine learning approaches also feature this fact, and strengthen it. Deep learning approaches have used the electron density both as input and output. On the one hand, machine learning has allowed to obtain electron densities for big and flexible systems with a very good scaling, which opens the way for having electron density analyses on big systems. On the other hand, machine learning models based on the electron density have been proven capable of providing a very effective connection to energetics, which will be a boost in functional development.

All in all, we can safely say, that the electron density is retrieving the central position it deserves.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

M.G. thankfully acknowledges the Spanish “Ministerio de Ciencia, Innovación y Universidades (MICIU)” for a predoctoral fellowship (FPU19/02903). M.G. and J.M. thank the Spanish "Ministerio de Ciencia en Innovación (MICINN)" for financial support, grant PGC2018-095953-B-I00; and the "Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y Tecnológica (FICyT)", grant IDI-2021-000054. B.L.-R. thanks SECTEI for the posdoctoral fellowship. J.M. acknoledges "Campus de Excelencia Internacional" in collaboration with Banco Santander in the framework of financial support for research mobility of the University of Oviedo.

Notes and references

1 T. Berlin, *J. Chem. Phys.*, 1951, **19**, 208–213.
2 H. Silberbach, *J. Chem. Phys.*, 1991, **94**, 2977–2985.
3 J. Hinze, *J. Chem. Phys.*, 1994, **101**, 6369–6370.
4 J. F. Rico, R. López, I. Ema and G. Ramírez, *J. Chem. Phys.*, 2002, **116**, 1788–1799.
5 Á. Nagy and C. Amovilli, *J. Chem. Phys.*, 2008, **128**, 114115.
6 S. Shahbazian, *Chem. - Eur. J.*, 2018, **24**, 5401–5405.
7 R. F. W. Bader, *Acc. Chem. Res.*, 1985, **18**, 9–15.
8 R. F. W. Bader, *Chem. Rev.*, 1991, **91**, 893–928.
9 *The Quantum Theory of Atoms in Molecules*, ed. C. F. Matta and R. J. Boyd, Wiley, 2007.
10 R. Laplaza, J. Munárriz and J. Contreras-García, *Conceptual Density Functional Theory: Towards a New Chemical Reactivity Theory*, Wiley-VCH GmbH, Weinheim, Germany, 2022, ch. Chemical Information.
11 A. M. Pendás, M. A. Blanco, A. Costales, P. M. Sánchez and V. Luaña, *Phys. Rev. Letters*, 1999, **83**, 1930–1933.
12 V. Luaña, P. Mori-Sánchez, A. Costales, M. A. Blanco and A. M. Pendás, *J. Chem. Phys.*, 2003, **119**, 6341–6350.
13 Q. K. Timerghazin and G. H. Peslherbe, *J. Chem. Phys.*, 2007, **127**, 064108.
14 R. Y. de Vries, W. J. Briels and D. Feil, *Phys. Rev. Letters*, 1996, **77**, 1719–1722.
15 L. A. Terrabuio, T. Q. Teodoro, M. G. Rachid and R. L. A. Haiduke, *J. Phys. Chem. A*, 2013, **117**, 10489–10496.

16 J. S. Anderson, L. Massa and C. F. Matta, *Chem. Phys. Lett.*, 2021, **780**, 138940.
17 J. Munárriz, F. Rabuffetti and J. Contreras-García, *Cryst. Growth Des.*, 2018, **18**, 6901–6910.
18 D. Cremer and E. Kraka, *J. Am. Chem. Soc.*, 1985, **107**, 3800–3810.
19 S. M. Bachrach, *Reviews in Computational Chemistry*, John Wiley & Sons, Inc., 2007, pp. 171–228.
20 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297.
21 T. A. Keith, *AIMALL (version 16.10.09)*, `http://aim.tkgristmill.com`, year = 2016.
22 F. Heidarzadeh and S. Shahbazian, *Int. J. Quantum Chem.*, 2010, **111**, 2788–2801.
23 F. Heidar-Zadeh, P. W. Ayers, T. Verstraelen, I. Vinogradov, E. Vöhringer-Martinez and P. Bultinck, *J. Phys. Chem. A*, 2017, **122**, 4219–4245.
24 R. Laplaza, F. Peccati, R. A. Boto, C. Quan, A. Carbone, J.-P. Piquemal, Y. Maday and J. Contreras-García, *WIREs Computational Molecular Science*, 2020, **11**, e1497.
25 E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *Journal of the American Chemical Society*, 2010, **132**, 6498–6506.
26 R. F. W. Bader and H. Essén, *J. Chem. Phys.*, 1984, **80**, 1943–1960.
27 R. F. W. Bader, R. J. Gillespie and P. J. MacDougall, *J. Am. Chem. Soc.*, 1988, **110**, 7329–7336.
28 S. Liu, C. Rong, T. Lu and H. Hu, *J. Phys. Chem. A*, 2018, **122**, 3087–3095.
29 S. Zhong, X. He, S. Liu, B. Wang, T. Lu, C. Rong and S. Liu, *J. Phys. Chem. A*, 2022, **126**, 2437–2444.
30 R. G. Parr, R. A. Donnelly, M. Levy and W. E. Palke, *J. Chem. Phys.*, 1978, **68**, 3801–3807.
31 P. Geerlings, F. D. Proft and W. Langenaeker, *Chem. Rev.*, 2003, **103**, 1793–1874.
32 R. G. Parr and R. G. Pearson, *J. Am. Chem. Soc.*, 1983, **105**, 7512–7516.
33 R. G. Parr and W. Yang, *J. Am. Chem. Soc.*, 1984, **106**, 4049–4050.
34 S. Liu and X. Zhang, *Acta Phys. Chim. Sin.*, 2018, **34**, 563–566.
35 W. F. Sanjuan-Szklarz, M. Woińska, S. Domagała, P. M. Dominiak, S. Grabowsky, D. Jayatilaka, M. Gutmann and K. Woźniak, *IUCrJ*, 2020, **7**, 920–933.
36 N. K. Hansen and P. Coppens, *Acta Crystallogr., Sect. A*, 1978, **34**, 909–921.
37 A. Fischer, D. Tiana, W. Scherer, K. Batke, G. Eickerling, H. Svendsen, N. Bindzus and B. Iversen, *J. Phys. Chem. A*, 2011, **115**, 13061–13071.
38 A. Volkov, C. Gatti, Y. Abramov and P. Coppens, *Acta Crystallogr., Sect. A*, 2000, **56**, 252–258.
39 N. Pérès, A. Boukhris, M. Souhassou, G. Gavoille and C. Lecomte, *Acta Crystallogr., Sect. B*, 1999, **55**, 1038–1048.
40 C. B. Hübschle, C. Ruhmlieb, A. Burkhardt, S. van Smaalen and B. Dittrich, *Zeitschrift für Kristallographie-Crystalline Materials*, 2018, **233**, 695–706.
41 A. Ø. Madsen, in *Modeling and Analysis of Hydrogen Atoms*, ed. D. Stalke, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 21–52.
42 L. J. Farrugia, *IUCrJ*, 2014, **1**, 265–266.
43 G. Cybenko, *Math. Control Signal Systems*, 1989, **2**, 303–314.
44 K. Hornik, M. Stinchcombe and H. White, *Neural Networks*, 1989, **2**, 359–366.
45 M. Gallegos, J. M. Guevara-Vela and A. M. Pendás, *J. Chem. Phys.*, 2022, **156**, 014112.
46 M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 058301.
47 S. Vargas, M. R. Hennefarth, Z. Liu and A. N. Alexandrova, *J. Chem. Theory Comput.*, 2021, **17**, 6203–6213.
48 D. C. Elton, Z. Boukouvalas, M. S. Butrico, M. D. Fuge and P. W. Chung, *Sci. Rep.*, 2018, **8**, 9059.
49 R. Zubatyuk, J. S. Smith, J. Leszczynski and O. Isayev, *Sci. Adv.*, 2019, **5**, eaav6490.
50 M. Hellström, M. Ceriotti and J. Behler, *J. Phys. Chem. B*, 2018, **122**, 10158–10171.
51 J. Behler, *J. Chem. Phys.*, 2016, **145**, 170901.
52 T. L. Fletcher and P. L. A. Popelier, *J. Comput. Chem.*, 2017, **38**, 1005–1014.
53 F. Guo, Y.-S. Wen, S.-Q. Feng, X.-D. Li, H.-S. Li, S.-X. Cui, Z.-R. Zhang, H.-Q. Hu, G.-Q. Zhang and X.-L. Cheng, *Comput. Mater. Sci.*, 2020, **172**, 109393.
54 P. Yoo, M. Sakano, S. Desai, M. M. Islam, P. Liao and A. Strachan, *npj Comput. Mater.*, 2021, **7**, 1–10.
55 J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192–3203.
56 L.-Y. Xue, F. Guo, Y.-S. Wen, S.-Q. Feng, X.-N. Huang, L. Guo, H.-S. Li, S.-X. Cui, G.-Q. Zhang and Q.-L. Wang, *Phys. Chem. Chem. Phys.*, 2021, **23**, 19457–19464.
57 O. A. von Lilienfeld, K.-R. Müller and A. Tkatchenko, *Nat. Rev. Chem*, 2020, **4**, 347–358.
58 A. Tkatchenko, *Nat. Rev. Chem*, 2020, **11**, 4895.
59 J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev and A. E. Roitberg, *J. Chem. Phys.*, 2018, **148**, 241733.
60 A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf and M. Ceriotti, *ACS Cent. Sci.*, 2018, **5**, 57–64.
61 A. Fabrizio, A. Grisafi, B. Meyer, M. Ceriotti and C. Corminboeuf, *Chem. Sci.*, 2019, **10**, 9424–9432.
62 F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke and K.-R. Müller, *Nat. Rev. Chem*, 2017, **8**, 872.
63 K. R. Briling, A. Fabrizio and C. Corminboeuf, *J. Chem. Phys.*, 2021, **155**, 024107.
64 M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew and K. A. Lyssenko, *Science*, 2017, **355**, 49–52.
65 P. Verma and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2017, **19**, 12898–12912.
66 T. Gould, *J. Chem. Theory Comput.*, 2017, **13**, 2373–2377.

67 D. S. Ranasinghe, A. Perera and R. J. Bartlett, *J. Chem. Phys.*, 2017, **147**, 204103.

68 R. Laplaza, V. Polo and J. Contreras-García, *Phys. Chem. Chem. Phys.*, 2019, **21**, 20927–20938.

69 K. P. Kepp, *Science*, 2017, **356**, 496.2–496.

70 K. Martin, *Angew. Chem., Int. Ed.*, 2017, **56**, 5396–5398.

71 É. Brémond, V. Tognetti, H. Chermette, J. C. Sancho-García, L. Joubert and C. Adamo, *J. Chem. Theory Comput.*, 2021, **18**, 293–308.

72 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Letters*, 1996, **77**, 3865–3868.

73 A. Zupan, K. Burke, M. Ernzerhof and J. P. Perdew, *J. Chem. Phys.*, 1997, **106**, 10184–10193.

74 J. Sun, A. Ruzsinszky and J. P. Perdew, *Phys. Rev. Lett.*, 2015, **115**, 036402.

75 J. Sun, R. C. Remsing, Y. Zhang, Z. Sun, A. Ruzsinszky, H. Peng, Z. Yang, A. Paul, U. Waghmare, X. Wu, M. L. Klein and J. P. Perdew, *Nat. Chem.*, 2016, **8**, 831–836.

76 M.-C. Kim, E. Sim and K. Burke, *Phys. Rev. Lett.*, 2013, **111**, 073003.

77 A. Wasserman, J. Nafziger, K. Jiang, M.-C. Kim, E. Sim and K. Burke, *Annu. Rev. Phys. Chem.*, 2017, **68**, 555–581.

78 S. Vuckovic, S. Song, J. Kozlowski, E. Sim and K. Burke, *J. Chem. Theory Comput.*, 2019, **15**, 6636–6646.

79 G. Santra and J. M. Martin, *J. Chem. Theory Comput.*, 2021, **17**, 1368–1379.

80 S. Song, S. Vuckovic, E. Sim and K. Burke, *J. Chem. Theory Comput.*, 2022, **18**, 817–827.

81 I. Mayer, I. Pápai, I. Bakó and Á. Nagy, *J. Chem. Theory Comput.*, 2017, **13**, 3961–3963.

82 S. Vuckovic and K. Burke, *J. Phys. Chem. Lett.*, 2020, **11**, 9957–9964.

83 A. Savin, R. Nesper, S. Wengert and T. F. Fässler, *Angew. Chem., Int. Ed. Engl.*, 1997, **36**, 1808–1832.

84 B. Silvi and A. Savin, *Nature*, 1994, **371**, 683–686.

85 P. Mori-Sánchez, A. J. Cohen and W. Yang, *Phys. Rev. Lett.*, 2008, **100**, 146401.

86 F. Peccati, R. Laplaza and J. Contreras-García, *J. Phys. Chem. C*, 2019, **123**, 4767–4772.

87 S. C. Capelli, H.-B. Bürgi, B. Dittrich, S. Grabowsky and D. Jayatilaka, *IUCrJ*, 2014, **1**, 361–379.

88 M. Woińska, S. Grabowsky, P. M. Dominiak, K. Woźniak and D. Jayatilaka, *Sci. Adv.*, 2016, **2**, e1600192.

89 M. Woińska, M. L. Chodkiewicz and K. Woźniak, *Chem. Commun.*, 2021, **57**, 3652–3655.

90 B. Landeros-Rivera, V. Jancik, R. Moreno-Esparza, D. Martínez Otero and J. Hernández-Trujillo, *Chem. - Eur. J.*, 2021, **27**, 11912–11918.

91 D. Jayatilaka, *Phys. Rev. Lett.*, 1998, **80**, 798.

92 D. Jayatilaka and D. J. Grimwood, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2001, **57**, 76–86.

93 S. Grabowsky, P. Luger, J. Buschmann, T. Schneider, T. Schirmeister, A. N. Sobolev and D. Jayatilaka, *Angew. Chem., Int. Ed.*, 2012, **51**, 6776–6779.

94 M. Woińska, D. Jayatilaka, B. Dittrich, R. Flaig, P. Luger, K. Woźniak, P. M. Dominiak and S. Grabowsky, *ChemPhysChem*, 2017, **18**, 3334–3351.

95 I. Bytheway, D. J. Grimwood, B. N. Figgis, G. S. Chandler and D. Jayatilaka, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2002, **58**, 244–251.

96 A. Genoni, L. H. Dos Santos, B. Meyer and P. Macchi, *IUCrJ*, 2017, **4**, 136–146.

97 D. Ramírez-Palma, B. Landeros-Rivera, A. Genoni, F. Cortés-Guzmán and J. Contreras-García, *Chemistry–Methods*, 2022, **2**, e20210004.

98 S. Grabowsky, A. Genoni, S. P. Thomas and D. Jayatilaka, in *The Advent of Quantum Crystallography: Form and Structure Factors from Quantum Mechanics for Advanced Structure Refinement and Wavefunction Fitting*, ed. D. Mingos and P. R. Raithby, Springer International Publishing, Cham, 2020, pp. 65–144.

99 B. Landeros-Rivera, J. Contreras-García and P. M. Dominiak, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2021, **77**, 715–727.

100 S. Raghunathan and U. D. Priyakumar, *Int. J. Quantum Chem.*, **n/a**, e26870.

101 L. David, A. Thakkar, R. Mercado and O. Engkvist, *J. Cheminf.*, 2020, **12**, 56.

102 J. Barker, J. Bulin, J. Hamaekers and S. Mathias, *LC-GAP: Localized Coulomb Descriptors for the Gaussian Approximation Potential*, Springer International Publishing, Cham, 2017, pp. 25–42.

103 J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.

104 J. Behler, *J. Chem. Phys.*, 2011, **134**, 074106.

105 M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi and P. Marquetand, *J. Chem. Phys.*, 2018, **148**, 241709.

106 M. P. Bircher, A. Singraber and C. Dellago, *Improved Description of Atomic Environments using Low-cost Polynomial Functions with Compact Support*, 2021.

107 K. Zhang, L. Yin and G. Liu, *Comput. Mater. Sci.*, 2021, **186**, 110071.

108 V. Zaverkin and J. Kästner, *J. Chem. Theory Comput.*, 2020, **16**, 5410–5421.

109 J. L. McDonagh, A. F. Silva, M. A. Vincent and P. L. A. Popelier, *J. Chem. Theory Comput.*, 2017, **14**, 216–224.

110 T. Fletcher, S. Kandathil and P. Popelier, *Theor. Chem. Acc.*, 2014, **133**, 1499–1459.

111 P. Maxwell, N. di Pasquale, S. Cardamone and P. L. A. Popelier, *Theor. Chem. Acc.*, 2016, **135**, 1–19.

112 P. L. A. Popelier, *Phys. Scr.*, 2016, **91**, 033007.

113 J. C. R. Thacker, A. L. Wilson, Z. E. Hughes, M. J. Burn, P. I. Maxwell and P. L. A. Popelier, *Mol. Simul.*, 2018, **44**, 881–890.

114 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. von Lilienfeld, *J. Chem. Theory Comput.*, 2015, **11**, 2087–2096.

115 B. Kalita, L. Li, R. J. McCarty and K. Burke, *Acc. Chem. Res.*, 2021, **54**, 818–826.

116 P. B. Jorgensen and A. Bhowmik, *DeepDFT: Neural Message Passing Network for Accurate Charge Density Prediction*, 2020.

117 J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. R. Castellanos, S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-Sánchez, D. Hassabis and A. J. Cohen, *Science*, 2021, **374**, 1385–1389.

118 P. Geerlings, E. Chamorro, P. K. Chattaraj, F. De Proft, J. L. Gázquez, S. Liu, C. Morell, A. Toro-Labbé, A. Vela and P. Ayers, *Theo. Chem. Acc.*, 2020, **139**, 1–18.

119 C. Rong, B. Wang, D. Zhao and S. Liu, *WIREs Comput. Mol. Sci.*, 2020, **10**, e1461.

120 R. G. Parr and Y. Weitao, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, 1995.