doi:10.2533/chimia.2020.232 Chimia 74 (2020) 232–236 © A. Fabrizio, K. Briling, A. Grisafi, C. Corminboeuf

Learning (from) the Electron Density: Transferability, Conformational and Chemical Diversity

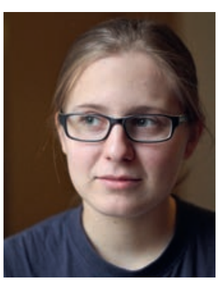
Alberto Fabrizio[§a], Ksenia Briling[a], Andrea Grisafi[b], and Clémence Corminboeuf*[a]

[§]SCS-Metrohm Award for the Best Oral Presentation in Computational Chemistry

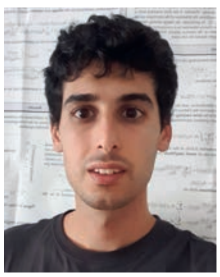
Abstract: Machine-learning in quantum chemistry is currently booming, with reported applications spanning all molecular properties from simple atomization energies to complex mathematical objects such as the many-body wavefunction. Due to its central role in density functional theory, the electron density is a particularly compelling target for non-linear regression. Nevertheless, the scalability and the transferability of the existing machine-learning models of $\rho(\mathbf{r})$ are limited by its complex rotational symmetries. Recently, in collaboration with Ceriotti and coworkers, we combined an efficient electron density decomposition scheme with a local regression framework based on symmetry-adapted Gaussian process regression able to accurately describe the covariance of the electron density spherical tensor components. The learning exercise is performed on local environments, allowing high transferability and linear-scaling of the prediction with respect to the number of atoms. Here, we review the main characteristics of the model and show its predictive power in a series of applications. The scalability and transferability of the trained model are demonstrated through the prediction of the electron density of Ubiquitin.

Keywords: Computational chemistry · Electron density · Machine learning · Quantum chemistry

Born in Italy, ***Alberto Fabrizio*** received his BSc and MSc in chemistry from the École Polytechnique Fédérale de Lausanne (EPFL). In 2016, he joined the group of Prof. Clémence Corminboeuf at École Polytechnique Fédérale de Lausanne (EPFL) as a PhD student. His current research focuses on characterizing and minimizing the errors of density functional approximations in new chemical situations and exploring the applicability of machine-learning models to access the density information.



Ksenia Briling received her Specialist degree in chemistry from Moscow State University, where she focused on the development of simple models for electronic structure theory. In 2019, she joined the Laboratory for Computational Molecular Design (LCMD) under the supervision of Prof. Clemence Corminboeuf at the EPFL as a PhD student. Currently, her research focuses on machine learning applications to solve quantum chemical problems.



Andrea Grisafi was born in Lucca (Italy) in 1991. He received his MSc in chemistry both from the University of Pisa and Scuola Normale Superiore of Pisa. In 2016, he joined the group of Michele Ceriotti at EPFL as a PhD student. His current research focuses on developing atomistic representations that are suitable to incorporate physical symmetries and long-range effects within machine-learning models of molecule and material properties

Clémence Corminboeuf started her independent career as a tenure track assistant professor at the EPFL in 2007. She was promoted to associate and then full professor in 2014 and 2019. In 2010, she received the silver medal at the European Young Chemist Award. She was awarded two ERC grants (starting and consolidator), the Werner Prize of the Swiss Chemical Society in 2014 and the Theoretical Chemistry Award from the ACS Physical Chemistry Division in 2018. She co-authored more than 150 publications focusing on the development of electronic structure methods as well as conceptual tools targeted for applications in the fields of organic electronics and homogeneous catalysis. Picture credit Alan Herzog, EPFL

*Correspondence: Prof. Dr. C. Corminboeuf, E-mail: clemence.corminboeuf@epfl.ch
[a]Laboratory for Computational Molecular Design, Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland; [b]Laboratory of Computational Science and Modeling, Institute of Material Science and Engineering, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

1. Introduction

Since the formulation of the Hohenberg-Kohn theorems,[1] the electron density [ρ(**r**)] has become an essential quantity to compute the electronic properties of atoms, molecules, and materials.[2] Its simple dependence on real space coordinates and the fact that it encodes the same information as the many-body wavefunction make ρ(**r**) a very appealing object. The common procedure to obtain the electron density for a given compound is to perform *ab initio* computations either within the framework of Kohn-Sham density functional theory (KS-DFT)[3] or with wavefunction based methods (Hartree-Fock, post-HF, *etc.*). Within this procedure, ρ(**r**) is variationally optimized for a given Hamiltonian or density functional approximation. Computing *ab initio* electron density is inevitably associated with the problem of scalability,[4a] *i.e.* the computational cost rises dramatically when targeting thousands of compounds or large and chemically diverse systems.

The scalability problem has a long history in quantum chemistry and has been addressed with the introduction of linear scaling techniques, including the divide-and-conquer approach originally proposed by Yang,[4] Mezey's molecular electron density LEGO assembler (MEDLA)[5] and the adjustable density matrix assembler (ADMA),[6] the fragment molecular orbital (FMO) approach of Kitaura and coworkers,[7] Karle's kernel energy methods (KEM)[8] and the molecular fractionation with conjugate caps (MFCC) proposed by Zhang.[9] Along with these techniques, the linear scaling reconstruction of the electron density has been proposed using extremely localized molecular orbitals, such as in the ALMO[10] and ELMO[10a,11] approaches. Given that ρ(**r**) is an experimental observable, X-ray diffraction is yet another technique to extract this information, usually in combination with multipolar models to reconstruct the data.[12] The latter exploits molecular fragment and pseudo-atomic libraries such as the experimental library of multipolar atom model (ELMAM),[13] its extension (ELMAM2),[14] the University of Buffalo database (UBDB),[15] the Invarioms set,[16] and the supramolecular synthon-based fragment approach (SBFA).[17]

Machine-learning models have been recently developed as an alternative cost-efficient technique to access the electron density information. The first model proposed by Tuckerman, Burke, Müller *et al.*[18] exploits the Hohenberg-Kohn mapping between the nuclear positions and the electron density. In particular, their method characterizes a chemical compound through a smoothed representation of its external potential and decomposes the electron density into an orthogonal basis set expansion. The method has been successfully applied to regress the density and drive the dynamics of small molecules (*i.e.* H_2, water, ethane, benzene and malonaldehyde). Nevertheless, the choice of an orthogonal expansion combined with a global representation (*i.e.* each molecule is represented as a whole indivisible entity, as opposed to a collection of atoms) somehow limits the transferability (*i.e.* predictions on larger systems, upon training on small molecules) of this landmark model. An alternative procedure based on neural networks to predict the electron density has been proposed by Ramprasad *et al.*[19] The electron density and the local density of states are, in this case, represented on a real space grid mapped to the local atomic environment around each grid-point. The Ramprasad model is thus transferable, but its reliance on millions of grid-points limits its numerical tractability and practicality. A similar approach was also proposed by Alred and coworkers in the context of the analysis of sulfur crosslinked carbon nanotubes.[20] A third strategy, proposed by some of the present authors together with Ceriotti and collaborators,[21,22] combines the numerical efficiency of the basis set expansion with a transferable framework based on a local atomic representation. In this model, the density is decomposed into an atom-centered, non-orthogonal basis and learned using symmetry-adapted Gaussian process regression (SA-GPR),[23] in order to capture the complex covariances of the electron density field. The method has been successfully applied to reproduce the charge density of a conformationally diverse ensemble of hydrocarbons and further improved to handle the rich chemistry of amino acid side-chain dimers taken from the BioFragment Database (BFDb).[24] This work gives an overview of the latter model and of the broad spectrum of properties derivable from ρ(**r**). Emphasis is placed on Hirshfeld atomic charges,[25] molecular dipole moments, electrostatic potential mappings and characterization of (non)covalent bonding patterns with the DORI scalar field.[26] Exploiting the linear-scaling of the model, we also demonstrate the transferability of the framework by predicting the electron density of the Ubiquitin protein as a prototypical macromolecule.

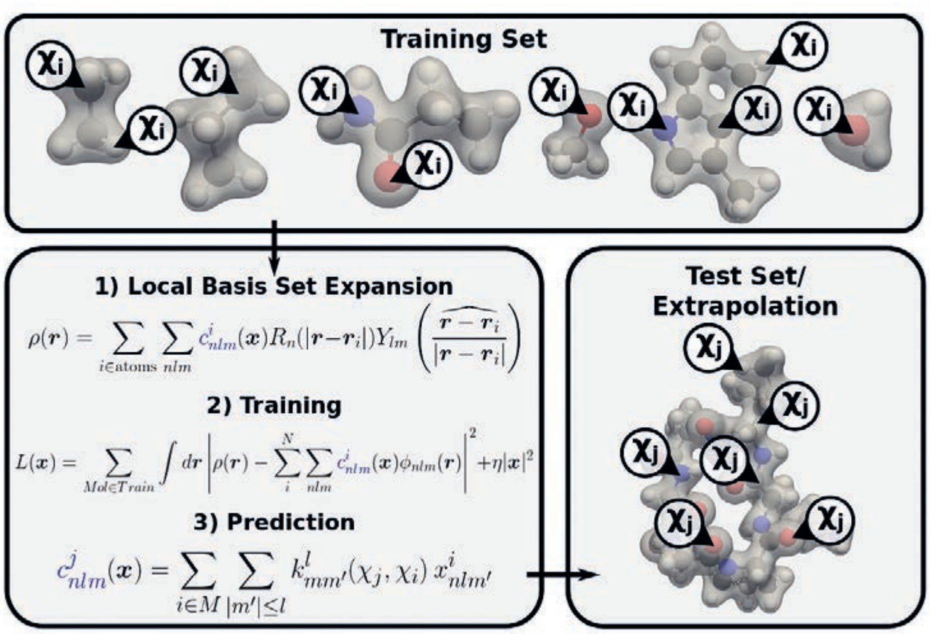


Fig. 1. Schematic representation of the regression framework for ρ(**r**). The electron densities of the training set are projected onto an atom-centred basis set, whose expansion coefficients (c^i_{nlm}, in blue) are the target of the prediction. The training is performed by finding the regression weights (**x**) that minimize a quadratic loss function [L(**x**)]. After training, the regression weights and the similarity kernels between the atomic environments of the reference (χ_i) and the unknown (χ_j) molecules are combined to obtain the predicted coefficients.

2. Overview of the Model

The SA-GPR model of ρ(**r**) relies on an atom-centered additive decomposition of the electron density (Eqn. (1)) inspired by other well-established practices in quantum chemistry, such as the divide-and-conquer[4] and fragment molecular orbitals (FMO) methods.[7] Here, the electron density is expanded onto an atom-centered non-orthogonal specialized basis set, commonly used in the context of the density fitting approximation (RI auxiliary basis).[27] These basis sets are specifically designed to closely mimic the features of one-electron densities (*i.e.* the binary product of atomic orbitals) and minimizes the density decomposition error.

$$\rho(\vec{r}) = \sum_{i}^{N_{\mathrm{atoms}}} \sum_{k}^{N_{\mathrm{RI}}} c_k^i \phi_{i,k}^{\mathrm{RI}}(\vec{r}) = \sum_{i}^{N_{\mathrm{atoms}}} \sum_{nlm} c_{nlm}^i R_n(r_i) Y_l^m(\theta,\phi) \qquad (1)$$

where the sums are performed over all the RI basis functions and over all atoms (i) in a molecule, k is a collective index for the radial (n) and angular channels (l, m), R_n are Gaussian radial functions and Y_l^m are spherical harmonics. The basis set expansion coefficients (c^i_{nlm}) are the final prediction target of the SA-GPR model and depend on the similarity kernels and the regression weights (see Fig. 1, Prediction). In this way, instead of choosing a particular scheme to decompose the density into atomic contributions *a priori* (*e.g.* Hirshfeld partitioning,[25] Edmiston-Ruedenberg scheme,[28] PDFT,[29] *etc.*), it is the machine-learning model that determines the best atomic decomposition by searching for the optimal atom-centered coefficients. During the training (Fig. 1, Training), the model aims at finding the regression weights (**x**) that minimize a quadratic loss function between *ab initio* density and its basis set decomposition. In the work reviewed here, the training set was composed by 2000 small organic dimers chosen randomly from the sidechain–sidechain interaction subset of the BFDb,[24] excluding sulfur-containing compounds. To predict the electron density of a new compound, the basis set coefficients (c^i_{nlm}) are computed combining the trained regression weights (**x**) with the similarity kernels between the atomic environments of the reference (χ_i) and the unknown (χ_j) molecule (Fig. 1). The rotational symmetries of the spherical tensor components of the density are encoded in the model using a reformulation of the GPR framework based on the SOAP representation[30] proposed by Ceriotti and coworkers (λ-SOAP). The λ-SOAP representation and similarity kernels have the characteristic to be covariant with the symmetry operations of the O(3) group.[23]

3. Applications

The applications reported in Fig. 2 illustrate the versatility and accuracy of the electron densities predicted with the ML model. In principle, any quantum chemical quantity can be extracted from ρ(**r**), but in practice the most common applications are the derivation of electrostatic moments (*e.g.* partial charges, dipoles, *etc.*), electrostatic potential mappings and, upon integration, electrostatic energies. Appealing applications of real-time access to the electron density as provided here are the evaluation of electrostatic potential maps and electrostatic complementarity between host–guest compounds (*e.g.* during the potential drug screening for a protein target)[31] or for the exact treatment of electrostatic interactions in molecular simulations. The upper panels of Fig. 2 show the dipole moments and the classical Hirshfeld dominant charges[25c] of a prototypical water molecule example as obtained from our ML densities. The formalism generally used to compute these two properties was modified to accommodate the fact that the predicted electron densities are expanded into a single set of orbitals (Eqn. (1)). In particular, the dipole moment can be written as

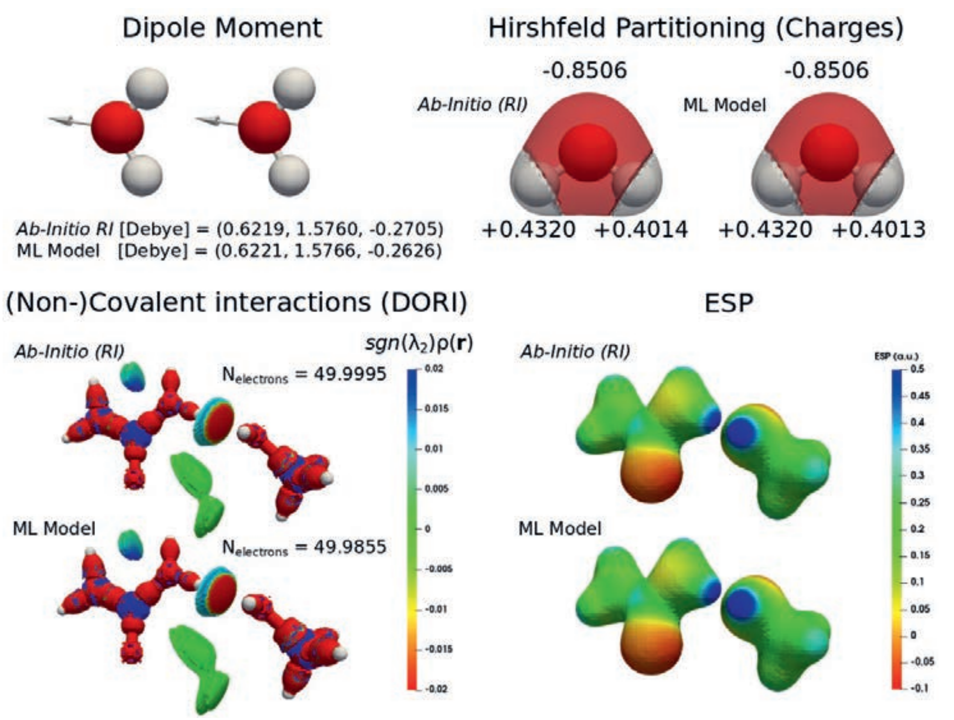


Fig. 2. Applications of the ML-electron density. (*top, left*) Molecular dipole moment (quantitative data reported in Debye). (*top, right*) Application of the classical Hirshfeld dominant partitioning (HCD) on ρ(**r**). The color code highlights the density regions belonging to the oxygen (red) and to each of the two hydrogens (white). HCD partial charges are also reported for each atom. The *ab initio* (RI) label refers to the density decomposed on the RI-basis set. (*bottom, left*) DORI map of the acetamide-methanol belonging to the BFDb (DORI isovalue: 0.09). The panel report additionally the number of electrons in this dimer, obtained from the integrated density. The isosurface is color-coded with sgn(λ_2) · ρ(**r**), which is a scalar field capable of identifying the regions dominated by covalent/strongly attractive interactions (red), steric clashes (blue) and weakly attractive forces (green). (*bottom, right*) Electrostatic potential map (ESP) of the same dimer, projected onto the 0.05 $e^{-}\cdot$Bohr^{-3} isosurface of the density. The color code distinguishes the regions with negative (red), intermediate (green) and positive (blue) ESP. Water densities were computed at PBE/cc-pVQZ, while the acetamide-methanol dimer was computed at ωB97X-D/cc-pVQZ.

$$\begin{aligned}\vec{\mu} &= \sum_{i}^{N_{\text{atoms}}} Q_i \vec{R}_i - \int \rho(\vec{r})\, \vec{r}\, d^3\vec{r} \\ &= \sum_{i}^{N_{\text{atoms}}} \vec{R}_i \left(Q_i - \sum_n c^i_{n00} \int R_n(r_i) Y^0_0(\theta,\phi)\, d^3\vec{r}_i \right) \\ &- \sum_{i}^{N_{\text{atoms}}} \sum_{nm} c^i_{n1m} \int R_n(r_i) Y^m_1(\theta,\phi)\, \vec{r}_i\, d^3\vec{r}_i\end{aligned} \tag{2}$$

where Q_i and $\vec{R_i}$ are the charge and position of the ith nucleus and $\vec{r_i} = \vec{r} - \vec{R_i}$. The first term depends on coefficients for s-functions and can be interpreted as the contribution of partial atomic charges. The second term, which contains coefficients for p-functions, takes into account higher-order effects. While the first term formally depends on nuclear positions, it does not depend on the choice of origin if the system is neutral, *i.e.*

$$\int \rho(\vec{r}) d^3\,\vec{r} = \sum_{i}^{N_{atoms}} Q_i.$$

Dipole moments are highly sensitive to small density variations and their ML-predictions are thus a very challenging task even for small molecules. The problem is further complicated by the fact that the model is not constrained to yield the exact number of electrons. In this respect, the accuracy of the predictions obtained for the water molecule is fairly impressive. The topological features of $\rho(\mathbf{r})$ are another relevant target commonly exploited for distinguishing different covalent bonding patterns and non-covalent interactions. The density overlap region indicator (DORI)[26] color-coded by the sign of the second density Hessian eigenvalue $[\text{sgn}(\lambda_2) \cdot \rho(\mathbf{r})]$[32,33] is given here as an example to reveal the dominant electrostatic (H-bond) nature of the interactions between the acetamide-methanol dimer (reference number 0557 in our database).[34] The excellent agreement between the *ab initio* and the ML-predicted DORI patterns, demonstrates that the predicted density has a correct topology (curvature) and an accurate amplitudes in each point of space capturing both the large amplitudes (red) typical of covalent and electrostatic attractive interactions and the density clashes in the low density regime (green) typical of the van der Waals regions. The excellent agreement is also observed between the ML-$\rho(\mathbf{r})$-based electrostatic potential maps and the DFT reference for the same dimer.

4. Transferability: The Electron Density of the Ubiquitin Protein

The key advantage of a local machine learning framework is that the learning exercise and, thus, the computation of the regression weights (**x** in Fig. 1) can be performed on any set of molecules, which becomes the starting point for predicting the targeted quantity on a much larger system. Within this framework, the 'extrapolation' from small to complex compounds is an interpolation between the local environments contained in the training set and those of the target molecule. Since the complexity of the prediction only depends on the number of atoms in the target, the regression of the electron density has a strictly linear scaling computational cost. In contrast, the *ab initio* computational complexity scales as the third power of the number of basis functions centered on each atom. In Fig. 3, we demonstrate the transferability of the SA-GPR model trained only on the small sidechain-sidechain dimers of the BFDb by predicting the electron density of the Ubiquitin protein (PDB ID: 1D3Z).

The machine-learning model allows accessing the density information of Ubiquitin at a pace incommensurably fast when compared to an equivalent quality *ab initio* computation (ωB97X-D/cc-pVQZ). The predicted electron density can be used in any of the applications mentioned in Section 4, among which we report as a general example the computation of the electrostatic potential maps.

5. Conclusions and Outlook

The electron density is a compelling target for machine-learning applications, as demonstrated by the broad spectrum of properties that can be derived upon prediction. Combining an efficient decomposition scheme and the SA-GPR framework,[23] the model is especially useful to target density applications in real space, such as the identification of covalent and non-covalent interaction fingerprints with scalar fields and electrostatic potential mappings. Due to its locality, the framework is highly transferable and can be successfully applied for the regression of complex or out-of-

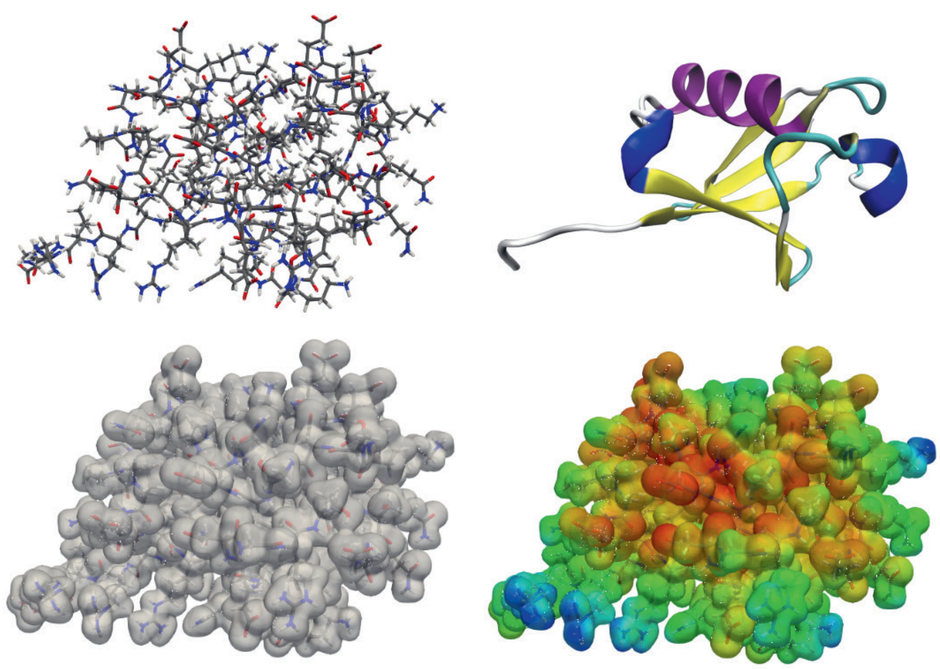


Fig. 3. (*top*) Structure of Ubiquitin [PDB ID: 1D3Z], using a skeletal model (*left*) and a ribbon diagram (*right*). The color code in the right panel highlights the different types of secondary structures. (*bottom, left*) Predicted electron density (isovalue: 0.01 $e^{-}\cdot$Bohr^{-3}). (*bottom, right*) Electrostatic potential map (ESP) computed from the predicted density. The color code distinguishes the regions with the most negative (red), intermediate (green) and most positive (blue) ESP. As no sulfur atoms are included in the original training set, the single methionine residue of the wild form of Ubiquitin has been substituted with a non-natural amino acid containing oxygen instead.

sample molecules, given a sufficiently diverse training set. The favorable linear scaling with the number of atoms allows accessing *ab initio* quality densities for extremely large systems, as shown in the case of the Ubiquitin protein. Overall, the model and the applications presented here represent a concrete example of a more comprehensive repertoire of novel methodologies, where kernel-based machine learning and artificial neural networks are applied to access quantum chemical properties. In perspective, the SA-GPR framework could be applied to the regression of any real space scalar field, as long as its decomposition in a Gaussian basis is rapidly convergent and it is covariant with spherical tensor transformations.

Acknowledgements

The work highlighted in this article has been supported by EPFL, by the National Centre of Competence in Research (NCCR) 'Materials' Revolution: Computational Design and Discovery of Novel Materials (MARVEL)' of the Swiss National Science Foundation (SNSF) and the European Research Council (ERC-CoG Grant 817977, PushQChem). The authors thank Michele Ceriotti for the collaboration. A.F. acknowledges Metrohm and the Swiss Chemical Society (SCS) for the SCS-Metrohm Best Oral Presentation Award.

Received: February 14, 2020

[1] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864, DOI: 10.1103/PhysRev.136.B864.

[2] a) R. G. Parr, W. Yang, 'Density-Functional Theory of Atoms and Molecules', Oxford University Press, New York, USA, **1989**; b) Fiolhais, F. Nogueira, M. A. L. Marques, 'A Primer in Density Functional Theory', Springer Berlin, Heidelberg, **2003**, DOI: 10.1007/3-540-37072-2; T. Tsuneda, 'Density Functional Theory in Quantum Chemistry', Springer Japan, Tokyo, **2014**, DOI: 10.1007/978-4-431-54825-6. C.

[3] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133, DOI: 10.1103/PhysRev.140.A1133.

[4] a) W. Yang, *Phys. Rev. Lett.* **1991**, *66*, 1438, DOI: 10.1103/PhysRevLett.66.1438; b) W. Yang, *Phys. Rev. A* **1991**, *44*, 7823, DOI: 10.1103/PhysRevA.44.7823; c) T. Zhu, W. Pan, W. Yang, *Phys. Rev. B* **1996**, *53*, 12713, DOI: 10.1103/PhysRevB.53.12713; d) W. Yang, T. Lee, *J. Chem. Phys.* **1995**, *103*, 5674, DOI: 10.1063/1.470549.

[5] a) P. D. Walker, P. G. Mezey, *J. Am. Chem. Soc.* **1993**, *115*, 12423, DOI: 10.1021/ja00079a025; b) P. D. Walker, P. G. Mezey, *J. Am. Chem. Soc.* **1994**, *116*, 12022, DOI: 10.1021/ja00105a050.

[6] a) T. E. Exner, P. G. Mezey, *J. Phys. Chem. A* **2002**, *106*, 11791, DOI: 10.1021/jp0263166; b) T. E. Exner, P. G. Mezey, *J. Comput. Chem.* **2003**, *24*, 1980, DOI: 10.1002/jcc.10340.

[7] a) K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, *Chem. Phys. Lett.* **1999**, *313*, 701, DOI: 10.1016/S0009-2614(99)00874-X; b) T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, K. Kitaura, *Chem. Phys. Lett.* **2000**, *318*, 614, DOI: 10.1016/S0009-2614(00)00070-1; c) K. Kitaura, S.-I. Sugiki, T. Nakano, Y. Komeiji, M. Uebayasi, *Chem. Phys. Lett.* **2001**, *336*, 163, DOI: 10.1016/S0009-2614(01)00099-9; d) Y. Inadomi, T. Nakano, K. Kitaura, U. Nagashima, *Chem. Phys. Lett.* **2002**, DOI: 10.1016/S0009-2614(02)01291-5; e) T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, K. Kitaura, *Chem. Phys. Lett.* **2002**, *351*, 475, DOI: 10.1016/S0009-2614(01)01416-6.

[8] a) L. Huang, L. Massa, J. Karle, *Int. J. Quantum Chem.* **2005**, *103*, 808, DOI: 10.1002/qua.20542; b) L. Huang, L. Massa, J. Karle, *Int. J. Quantum Chem.* **2006**, *106*, 447, DOI: 10.1002/qua.20760; c) L. Huang, L. Massa, J. Karle, *Proc. Natl. Acad. Sci.* **2006**, *103*, 1233, DOI: 10.1073/pnas.0510342103; d) L. Huang, L. Massa, J. Karle, *Proc. Natl. Acad. Sci.* **2008**, *105*, 1849, DOI: 10.1073/pnas.0711297105.

[9] a) D. W. Zhang, J. Z. H. Zhang, *J. Chem. Phys.* **2003**, *119*, 3599, DOI: 10.1063/1.1591727; b) A. M. Gao, D. W. Zhang, J. Z. H. Zhang, Y. Zhang, *Chem. Phys. Lett.* **2004**, *394*, 293, DOI: 10.1016/j.cplett.2004.06.137; c) X. He, J. Z. H. Zhang, *J. Chem. Phys.* **2006**, *124*, 184703, DOI: 10.1063/1.2194535; d) X. He, J. Z. H. Zhang, *J. Chem. Phys*. **2005**, *122*, 031103, DOI: 10.1063/1.1849132.

[10] a) H. Stoll, G. Wagenblast, H. Preuß, *Theor. Chim. Acta* **1980**, *57*, 169, DOI: 10.1007/BF00574903; b) R. Z. Khaliullin, M. Head-Gordon, A. T. Bell, *J. Chem. Phys.* **2006**, *124*, 204105, DOI: 10.1063/1.2191500.

[11] a) M. Sironi, A. Genoni, M. Civera, S. Pieraccini, M. Ghitti, *Theor. Chem. Acc.* **2007**, *117*, 685, DOI: 10.1007/s00214-006-0200-7; b) B. Meyer, B. Guillot, M. F. Ruiz-Lopez, A. Genoni, *J. Chem. Theory Comput.* **2016**, *12*, 1052, DOI: 10.1021/acs.jctc.5b01007.

[12] a) F. L. Hirshfeld, *Acta Crystallogr. Sect. B* **1971**, *27*, 769, DOI: 10.1107/S0567740871002905; b) R. F. Stewart, *Acta Crystallogr. Sect. A* **1976**, *32*, 565, DOI: 10.1107/S056773947600123X; c) N. K. Hansen, P. Coppens, *Acta Crystallogr. Sect. A* **1978**, *34*, 909, DOI: 10.1107/S0567739478001886.

[13] a) V. Pichon-Pesme, C. Lecomte, H. Lachekar, *J. Phys. Chem.* **1995**, *99*, 6242, DOI: 10.1021/j100016a071; b) C. Jelsch, V. Pichon-Pesme, C. Lecomte, A. Aubry, *Acta Crystallogr. Sect. D* **1998**, *54*, 1306, DOI: 10.1107/S0907444998004466; c) B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte, C. Jelsch, *Acta Crystallogr. Sect. A* **2007**, *63*, 108, DOI: 10.1107/S0108767306053748; d) C. Lecomte, C. Jelsch, B. Guillot, B. Fournier, A. Lagoutte, *J. Synchrotron Radiat.* **2008**, *15*, 202, DOI: 10.1107/S0909049508000447.

[14] a) S. Domagała, P. Munshi, M. Ahmed, B. Guillot, C. Jelsch, *Acta Crystallogr. Sect. B* **2011**, *67*, 63, DOI: 10.1107/S0108768110041996; b) S. Domagała, B. Fournier, D. Liebschner, B. Guillot, C. Jelsch, *Acta Crystallogr. Sect. A* **2012**, *68*, 337, DOI: 10.1107/S0108767312008197.

[15] a) P. Kumar, M. K. Cabaj, P. M. Dominiak, *Crystals* **2019**, *9*, 668, DOI: 10.3390/cryst9120668; b) P. M. Dominiak, A. Volkov, X. Li, M. Messerschmidt, P. Coppens, *J. Chem. Theory Comput.* **2007**, *3*, 232, DOI: 10.1021/ct6001994.

[16] B. Dittrich, T. Koritsánszky, P. Luger, *Angew. Chem. Int. Ed.* **2004**, *43*, 2718, DOI: 10.1002/anie.200353596.

[17] V. R. Hathwar, T. S. Thakur, T. N. G. Row, G. R. Desiraju, *Cryst. Growth Des.* **2011**, *11*, 616, DOI: 10.1021/cg101540y.

[18] a) F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, K.-R. Müller, *Nat. Comm.* **2017**, *8*, 872, DOI: 10.1038/s41467-017-00839-3; b) M. Bogojeski, F. Brockherde, L. Vogt-Maranto, L. Li, M. E. Tuckerman, K. Burke, K.-R. Müller, *arXiv*:1811.06255 2018.

[19] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, R. Ramprasad, *npj Comput. Mater.* **2019**, *5*, 22, DOI: 10.1038/s41524-019-0162-7.

[20] J. M. Alred, K. V. Bets, Y. Xie, B. I. Yakobson, *Compos. Sci. Technol.* **2018**, *166*, 3, DOI: 10.1016/j.compscitech.2018.03.035.

[21] A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, M. Ceriotti, *ACS Cent. Sci.* **2019**, *5*, 57, DOI: 10.1021/acscentsci.8b00551.

[22] A. Fabrizio, A. Grisafi, B. Meyer, M. Ceriotti, C. Corminboeuf, *Chem. Sci.* **2019**, *10*, 9424, DOI: 10.1039/C9SC02696G.

[23] A. Grisafi, D. M. Wilkins, G. Csányi, M. Ceriotti, *Phys. Rev. Lett.* **2018**, *120*, 036002, DOI: 10.1103/PhysRevLett.120.036002.

[24] L. A. Burns, J. C. Faver, Z. Zheng, M. S. Marshall, D. G. A. Smith, K. Vanommeslaeghe, A. D. MacKerell, K. M. Merz, C. D. Sherrill, *J. Chem. Phys.* **2017**, *147*, 161727, DOI: 10.1063/1.5001028.

[25] a) F. L. Hirshfeld, *Theor. Chim. Acta* **1977**, *44*, 129, DOI: 10.1007/BF00549096; b) P. Bultinck, C. Van Alsenoy, P. W. Ayers, R. Carbó-Dorca, *J. Chem. Phys.* **2007**, *126*, 144111, DOI: 10.1063/1.2715563; c) P. Bultinck, D. L. Cooper, R. Ponec, *J. Phys. Chem. A* **2010**, *114*, 8754, DOI: 10.1021/jp101707w.

[26] P. de Silva, C. Corminboeuf, *J. Chem. Theory Comput.* **2014**, *10*, 3745, DOI: 10.1021/ct500490b.

[27] a) J. L. Whitten, *J. Chem. Phys.* **1973**, *58*, 4496, DOI: 10.1063/1.1679012; b) B. I. Dunlap, J. W. D. Connolly, J. R. Sabin, *Int. J. Quantum Chem. Symp.* **1977**, *11*, 81, DOI: 10.1002/qua.560120813; c) M. Feyereisen, G. Fitzgerald, A. Komornicki, *Chem. Phys. Lett.* **1993**, *208*, 359, DOI: 10.1016/0009-2614(93)87156-W; d) K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *240*, 283, DOI: 10.1016/0009-2614(95)00621-A; e) F. Weigend, *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285, DOI: 10.1039/b204199p.

[28] C. Edmiston, K. Ruedenberg, *Rev. Mod. Phys.* **1963**, *35*, 457, DOI: 10.1103/RevModPhys.35.457.

[29] a) P. Elliott, M. H. Cohen, A. Wasserman, K. Burke, *J. Chem. Theory Comput.* **2009**, *5*, 827, DOI: 10.1021/ct9000119; b) P. Elliott, K. Burke, M. H. Cohen, A. Wasserman, *Phys. Rev. A* **2010**, *82*, 024501, DOI: 10.1103/PhysRevA.82.024501.

[30] A. P. Bartók, R. Kondor, G. Csányi, *Phys. Rev. B* **2013**, *87*, 184115, DOI: 10.1103/PhysRevB.87.184115.

[31] M. R. Bauer, M. D. Mackey, *J. Med. Chem.* **2019**, *62*, 3036, DOI: 10.1021/acs.jmedchem.8b01925.

[32] a) R. F. W. Bader, *Chem. Rev.* **1991**, *91*, 893, DOI: 10.1021/cr00005a013; b) R. F. M. Bader, 'The Quantum Theory of Atoms in Molecules', Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2007**, DOI: 10.1002/9783527610709.

[33] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, W. Yang, *J. Am. Chem. Soc.* **2010**, *132*, 6498, DOI: 10.1021/ja100936w.

[34] MaterialsCloud, *https://archive.materialscloud.org/2019.0071/v1*

License and Terms