Learning on-top: regressing the on-top pair density for real-space visualization of electron correlation

Alberto Fabrizio,[1, 2, a)] Ksenia R. Briling,[1, a)] David D. Girardier,[1] and Clemence Corminboeuf[1, 2, b)]

[1)]*Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering,*
École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

[2)]*National Centre for Computational Design and Discovery of Novel Materials (MARVEL),*
École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

(Dated: 28 May 2021)

The on-top pair density [$\Pi(\mathbf{r})$] is a local quantum chemical property, which reflects the probability of two electrons of any spin to occupy the same position in space. Simplest quantity related to the two-particles density matrix, the on-top pair density is a powerful indicator of electron correlation effects and, as such, it has been extensively used to combine density functional theory and multireference wavefunction theory. The widespread application of $\Pi(\mathbf{r})$ is currently hindered by the need for post-Hartree–Fock or multireference computations for its accurate evaluation. In this work, we propose the construction of a machine learning model capable of predicting the CASSCF-quality on-top pair density of a molecule only from its structure and composition. Our model, trained on the GDB11-AD-3165 database, is able to predict with minimal error the on-top pair density of organic molecules bypassing completely the need for *ab-initio* computations. The accuracy of the regression is demonstrated using the on-top ratio as a visual metric of electron correlation effects and bond-breaking in real-space. In addition, we report the construction of a specialized basis set, built to fit the on-top pair density in a single, atom-centered expansion. This basis, cornerstone of the regression, could be potentially used also in the same spirit of the resolution-of-the-identity approximation for the electron density.

I. INTRODUCTION

The spinless on-top pair density [OTPD, $\Pi(\mathbf{r})$, Eq. 1] is a local quantum chemical quantity, which represents the probability density for two electrons of any spin to occupy the same position in real-space:[1]

$$\Pi(\mathbf{r}) = \binom{N}{2} \int |\Psi(\mathbf{x}_1, ...\mathbf{x}_N)|^2 \, d\sigma_1 ... d\sigma_N d\mathbf{r}_3 ... d\mathbf{r}_N|_{\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r}}, \tag{1}$$

where σ is the spin variable, $\mathbf{r}$ is the space variable and $\mathbf{x} = (\mathbf{r}, \sigma)$.

Because the on-top pair density is the simplest local property that is still related to the two-particle density matrix (2-PDM), $\Pi(\mathbf{r})$ occupies a prominent place in the history of electron correlation in density functional functional theory. Already in the mid-90s it was shown that the on-top pair density is the fundamental quantity that justifies the accurate results obtained on some multireference systems with broken-symmetry LSDA and GGA computations.[2] Indeed, chemical situations such as the stretched H_2 molecule suffer from an apparent dilemma (akin to Löwdin's symmetry dilemma in Hartree–Fock):[3] preserving the correct spin distribution yields inaccurate electronic energies, and viceversa.[4,5] Concrete examples of this problem have been already known since the late 70s, when it was demonstrated that the local spin-density

approximation (LSDA) is only justified for single-determinant states and that the local spin density philosophy is not compatible with the spin densities of multireference wavefunctions.[6] In contrast, Perdew, Ernzerhof, Savin and Burke showed that despite yielding incorrect spin distribution, broken-symmetry solutions already at the LSDA level mimicked very closely the on-top pair density computed with multireference wavefunction theory.[2,7] It was later argued that the ability of LSDA to produce satisfyingly accurate spin-independent quantities [$\rho(\mathbf{r})$ and $\Pi(\mathbf{r})$] is the reason beyond the unexpectedly fair behavior in many chemical situations of this otherwise very crude approximation to exact DFT.[8]

Besides its usefulness to re-interpret the results of approximate functionals, the on-top pair density has been traditionally used to combine wavefunction theory and DFT. As early as the 1991, the pioneering work of Moscardó and San-Fabián[9,10] showed that it is possible to use the on-top pair density from a configuration-interaction wavefunction (CID) to help density functionals to capture the intrinsic two-body nature of the correlation energy. This first result only slightly anticipated the work of Becke, Savin and Stoll who proposed a complete re-evaluation of the local spin-density approach and replaced the spin densities [$\rho_\alpha(\mathbf{r})$, $\rho_\beta(\mathbf{r})$] as independent variables with $\rho(\mathbf{r})$ and $\Pi(\mathbf{r})$.[11] Since the electron density and the on-top pair density can be evaluated straightforwardly from a multiconfigurational wavefunction, the proposed substitution allows extending the machinery of Kohn–Sham DFT to multi-determinant reference states.

On the momentum of these works, the on-top pair density has been then mostly applied as ingredient of mixed electronic structure methods such as multiconfiguration DFT

[a)]The authors contributed equally to this work.
[b)]Electronic mail: clemence.corminboeuf@epfl.ch

(MC-DFT),[12] complete active space DFT (CAS-DFT)[13–16] and, more recently, multiconfiguration pair-density functional theory (MC-PDFT).[17–19] In these latter works, at least part of the total electronic energy is obtained non-variationally through a density functional, whose input is or directly derives from the on-top pair density of a converged multiconfigurational wavefunction. Among these methods, MC-PDFT has emerged as one of the closest realizations of the original suggestion to combine multireference wavefunction theory and DFT through the on-top pair density.[11] In the last few years, this framework has been shown to produce benchmark-quality results for typically challenging electronic structures,[20] including charge-transfer complexes,[21] main group and transition metal thermochemistry,[22,23] barrier heights,[22] and electronic excitations.[24] Complementary to the applications of the on-top pair density for method development, the success of MC-PDFT fueled a renewed interest in its potential application as a visualization tool, especially in the form of the on-top ratio [*i.e.* the ratio between the $\Pi(\mathbf{r})$ and $\rho^2(\mathbf{r})/4$] and its partial derivatives.[25] These quantities are effective real-space metrics to identify and visualize the effects of electron correlation, as well as to characterize fundamentally the different types of bonds and covalent-bond breaking.[25,26]

While the applications of the on-top pair density demonstrate its key role in the quantitative and qualitative description of electron correlation, its wide-spread use is severely limited by the necessity to perform post-Hartree–Fock or, more commonly, multireference computations to obtain accurate two-particle density matrices. In the last few years, the recourse to artificial intelligence, more precisely to machine learning (ML), has been increasingly proposed as an effective strategy to access fundamental quantum chemical objects bypassing the bottleneck of computationally demanding methods.[27–35] In this context, two of us have recently proposed,[36] improved[37] and demonstrated the large applicability[37,38] of a local machine learning model of the molecular electron density capable of accurately reproducing the complex rotational symmetries of the density in real-space. While applied to date only for the regression of the ground-state electron density, the architecture of the framework is general and allows, in principles, to target any local field as long as its decomposition onto an atom-centered basis set is possible.

In this work, we adapt the electron density learning framework to build a direct mapping between the structure and composition of a molecule and its CASSCF-quality on-top pair density. Granting a ready access to the information contained in $\Pi(\mathbf{r})$, we conceive our model as an effective tool for the real-space visualization of electron correlation effects and covalent-bond breaking. To achieve this goal, the framework is built on the recently introduced GDB11-AD-3165 database of Kulik and coworkers,[39] which consists of 3165 small organic molecules both at their equilibrium and distorted geometries. The GDB11-AD-3165 set has not only the advantage of hosting inherently multiconfigurational molecules and structures, but it also reports accurate CASSCF results, including the weight of the dominant electronic configuration (C_0^2) and the precise extent of the active spaces used.

II. THE $\phi^{\text{OTPD}}(\mathbf{r})$ SPECIALIZED BASIS SET

The symmetry-adapted Gaussian process regression framework (SA-GPR) used in this work was originally proposed by Csányi and Ceriotti and coworkers[41] and is a generalization of traditional GPR, capable of encoding the complex symmetries of Cartesian and spherical tensors of any order through a hierarchy of kernels. As already demonstrated in the case of the ground-state electron density,[36,37] SA-GPR captures accurately the rotational symmetries of a local scalar field, subject to the constraint that its angular dependence is described by a single set of atom-centered spherical harmonics [$Y_l^m(\theta, \phi)$]. In this way, any field can be decomposed into a sum of spherical tensor components with rank l, beginning with the spherically symmetric elements ($l = 0$) and increasingly adding anisotropy ($l > 0$). The treatment of the radial part of the field is subordinated to a much lighter constraint, as the only requirement is for the chosen atom-centered basis functions to have a similar long-range ($r \to \infty$) asymptotic behavior.

The local decomposition strategy is not only an efficient mathematical expedient to represent a scalar field and ensure its compatibility with SA-GPR, but it is also the foundation of the scalability and transferability of the model. These two last properties stem from the combination of the locality of the decomposed field with the locality of the molecular representation used within the SA-GPR framework ($\lambda-$SOAP) and allow the linear-scaling prediction of the target property for a complex molecule, while restraining the training procedure only to much smaller fragments. This type of LEGO approach is very similar in spirit to well-established linear scaling techniques in quantum chemistry such as Mezey's molecular electron density LEGO assembler (MEDLA)[42,43] and adjustable density matrix assembler (ADMA).[44–46]

For these reasons, we seek a decomposition of the on-top pair density [$\Pi(\mathbf{r})$] in the form of

$$\Pi(\mathbf{r}) = \frac{1}{2} \sum_{abcd}^{N_{AO}} D_{abcd} \chi_a(\mathbf{r}) \chi_b(\mathbf{r}) \chi_c(\mathbf{r}) \chi_d(\mathbf{r}) = \sum_i c_i \phi_i(\mathbf{r}), \quad (2)$$

where D_{abcd} is the two-particle density matrix at any chosen level of theory (*e.g.* HF, DFT, CASSCF, CCSD(T), *etc.*), $\chi(\mathbf{r})$ are atomic orbitals, c_i are decomposition expansion coefficients [$i = \{\text{atom}, n, l, m\}$] and $\phi_i(\mathbf{r})$ are a set of suitable basis functions for the decomposition.

The expansion onto the non-orthogonal basis shown in rightmost side of Eq. 2 has to be interpreted in the same spirit as the well-established density-fitting approximation (RI or resolution-of-the-identity) in the context of the electron density. We have already demonstrated that the use of specialized basis sets (auxiliary or RI-basis) optimized for the density-fitting approximation improves dramatically the regression of the electron density within the symmetry-adapted framework. However, as reported for the show-case water molecule in the top, left panel of Figure 1, common RI-basis functions result in unacceptable errors in the decomposition the on-top pair density, introducing instabilities in the shape of spurious anisotropic features in the field.
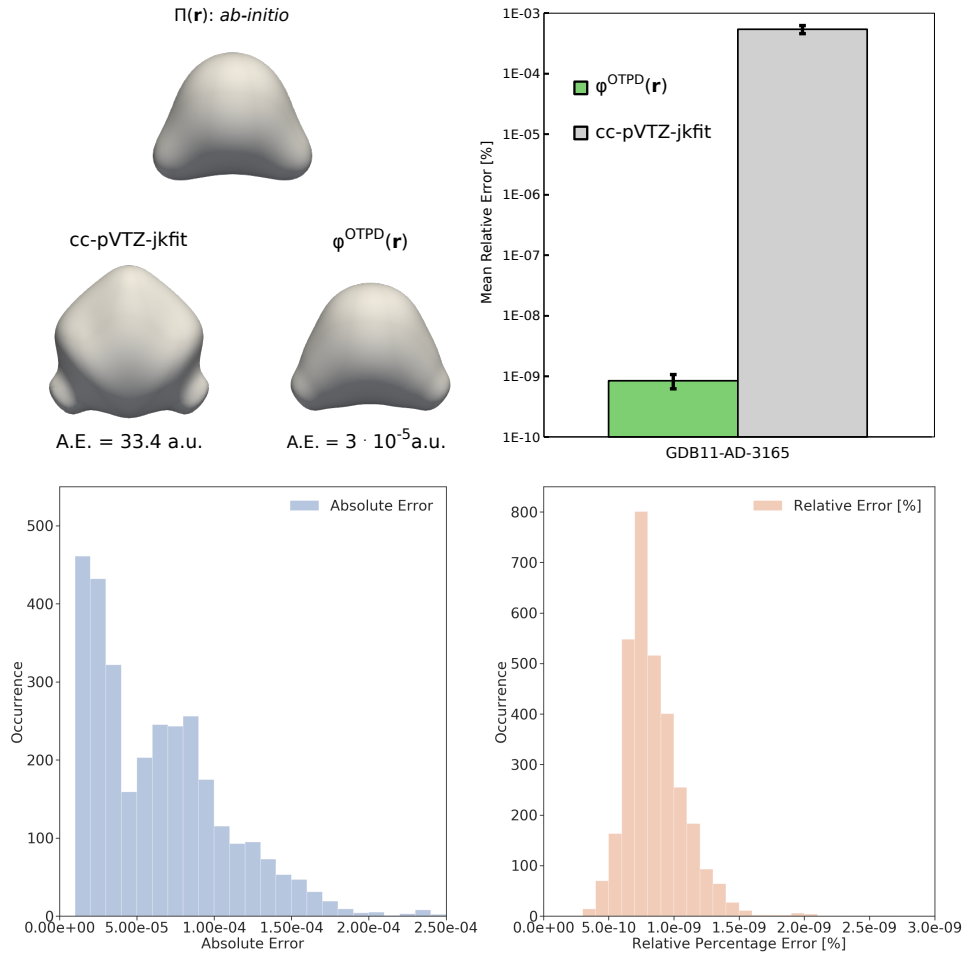
FIG. 1. *top, left*) Qualitative comparison of fitting performance on the water molecule; isosurfaces: $0.005 \text{ e}^2 \cdot \text{Bohr}^{-6}$. (*top, right*) Comparison between the $\phi^{OTPD}(\mathbf{r})$ and the cc-pVTZ-jkfit[40] (RI-basis) for the fitting of the on-top pair density (CASSCF/cc-pvTZ) on the GDB11-AD-3165 set.[39] Height of the histogram correspond to the mean relative percentage error on the GDB11-AD-3165. *bottom*) Distribution of the absolute and relative [%] fitting error for $\phi^{OTPD}(\mathbf{r})$ on the GDB11-AD-3165 set.

To better understand the reason behind the spectacular failure of RI-basis sets for the fitting of the on-top pair density, it is sufficient to consider the form of $\Pi(\mathbf{r})$ at the Hartree–Fock level:[11]

$$\Pi^{HF}(\mathbf{r}) = \frac{1}{2} \sum_{abcd}^{N_{AO}} [D_{ab}D_{cd} - \frac{1}{2}D_{ad}D_{cb}]\chi_a(\mathbf{r})\chi_b(\mathbf{r})\chi_c(\mathbf{r})\chi_d(\mathbf{r})$$
$$= \frac{\rho^2(\mathbf{r})}{4},$$

(3)

where D_{ab} is the Hartree–Fock one-particle density matrix and $\rho(\mathbf{r})$ the electron density. Equation 3 shows that, even excluding any effect due to electron correlation, the on-top pair density has higher amplitudes near the nuclei and decays much faster (twice as fast) than electron electron density, for which the RI-basis were optimized. Unfortunately, simple solutions such as doubling the exponents of the RI-basis improves only marginally the fitting error (A.E. = 5.8 a.u. on the

water molecule; standard RI-basis A.E. = 33.5 a.u.).

For this reason, we propose here the construction of a specialized basis [$\phi^{OTPD}(\mathbf{r})$], adapted to reduce the fitting error on the on-top pair density. Inspired by the work of Neese and coworkers,[47] we divide the optimization procedure in two distinct phases. A first set of basis exponents (α) for each angular momentum (up to $l = 4$) and for each element in the GDB11-AD-3165 set (H, C, N, O) is generated by optimization of the three coefficients (α_0, β, and γ) of the well-tempered formula:[48]

$$\alpha_i = \beta \, \alpha_{i-1} \left(1 + \gamma \left(\frac{i-2}{N+1}\right)^2\right), \quad i = 2, \ldots, N \quad (4)$$

The optimization is performed to minimize the fitting error of the on-top pair density at HF/cc-pVTZ of H_2, H_2O, H_2O_2, C_2H_2, C_2H_4, C_2H_6, C_6H_6, CH_4, NH_3, N_2H_2, and N_2 (the geometries were taken from the original set by Ahlrichs and coworkers).[49] This first crude optimization was then used as

an initial condition for the complete relaxation of each exponent with analytical gradients using the BFGS[50–53] algorithm. Further details about $\phi^{\text{OTPD}}(\mathbf{r})$ and its construction are reported in the Supplementary Material.

As the on-top pair density is a powerful indicator of static (or left-right,[54] or non-dynamical)[55] electron correlation, we choose to build our model on the recently published GDB11-AD-3165 dataset.[39] Figure 1 shows the comparison of the decomposition error on GDB11-AD-3165 between the optimized $\phi^{\text{OTPD}}(\mathbf{r})$ and the standard cc-pVTZ-jkfit[40] basis. The on-top pair densities were obtained at the CASSCF/cc-pVTZ level using the same active spaces as prescribed in the original reference.[39]

In the figure, the fitting error of $\phi^{\text{OTPD}}(\mathbf{r})$ is reported using two distinct integral metrics: the absolute error $\text{AE} = \int d\mathbf{r} \left| \Pi(\mathbf{r}) - \sum_i c_i \phi^{\text{fitting}}(\mathbf{r}) \right|^2$ and the relative error $\text{RE} = \int d\mathbf{r} \left| \Pi(\mathbf{r}) - \sum_i c_i \phi^{\text{fitting}}(\mathbf{r}) \right|^2 / \int d\mathbf{r} \Pi^2(\mathbf{r})$. The absolute error is extremely sensitive to the inaccuracies of the fitting near the nuclei and, generally, in the regions with high OTPD amplitudes. On the other hand, the relative error is a more fair metric both close and far from the nuclei and, in contrast to the absolute error, does not depend on the system size. Regardless of the metric chosen, however, $\phi^{\text{OTPD}}(\mathbf{r})$ is orders of magnitudes more accurate for the fitting of the on-top pair density than standard density-fitting basis sets of comparable size. From a chemical perspective, the GDB11-AD-3165 is divisible between compounds with marked multiconfigurational character (often distorted structures) and single-reference molecule (often at the equilibrium geometry). While this distinction exist in the dataset, the fitting error does not depend on the character of the molecular ground-state (see Supplementary Material for further details). Since the fitting error persists also after the learning procedure, reaching a high accuracy in the decomposition step is the cornerstone of any application of the machine learned on-top pair density.

III. LEARNING CURVES

The optimization of the specialized $\phi^{\text{OTPD}}(\mathbf{r})$ allows the efficient projection of the on-top pair density onto a non-orthogonal, atom-centered (local) basis and promotes its compatibility with the symmetry-adapted Gaussian process regression (SA-GPR) framework. To train the model and assess its accuracy, the GDB11-AD-3165 dataset was randomly split into a training set of 2550 molecules and a test set containing the remaining 615 compounds (∼20 % of the total). Because of its size and its chemical diversity, the GDB11-AD-3165 database contains a tremendous number of atom-centered environments ($\chi \sim$36'000). On the other hand, most of these local environments are very similar to each other and are redundant for the regression purposes. The locality of the molecular representation combined with the relative redundancy of chemistry at a short, atom-centered cutoff allows reducing significantly the computational cost of the regression by only choosing a subset of the M most different environments. Akin to our previous work on the electron density,

we set $M = 1000$ and perform a farthest point sampling in the environment space to form a basis for the regression of the on-top pair density.

The performance of the regression model, in terms of mean absolute and relative errors (as defined in Section II) is summarized in Figure 2. Besides the on-top pair density, the learning curve of the squared half-densities $[(\rho(\mathbf{r})/2)^2]$ are reported for comparison.
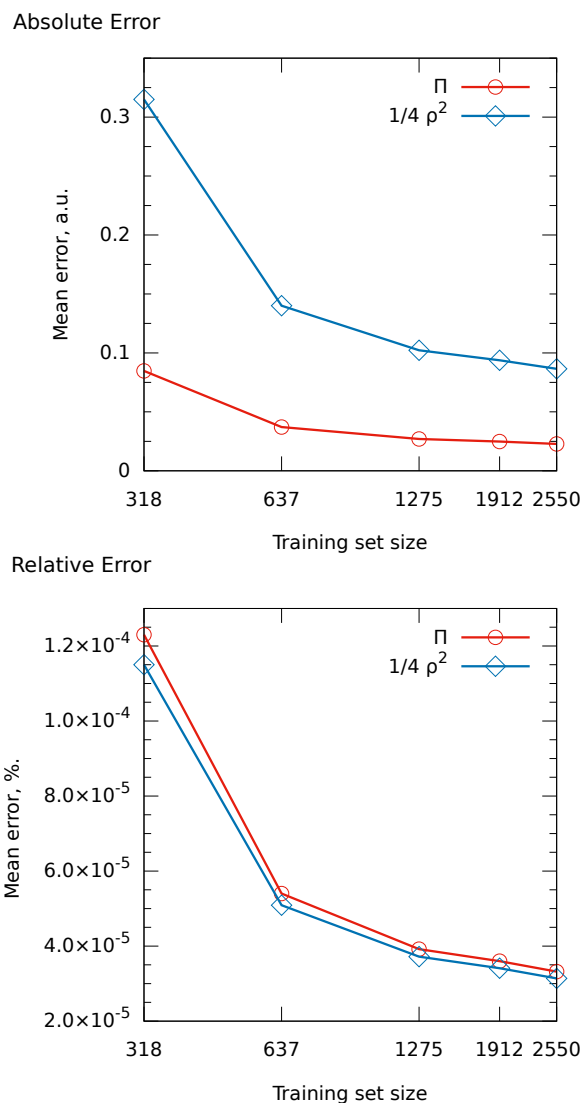

FIG. 2. Learning curves with respect to fitted on-top pair densities and squared half-densities (ML error). The learning curves of $(\rho(\mathbf{r})/2)^2$ are shown for comparison purposes. *(top)* Mean absolute error (MAE), *(bottom)* mean relative percentage error on the test set (615 molecules). The color code distinguishes the on-top pair density (red) from the squared half-densities (blue).

In the limit where the on-top pair density is computed at Hartree–Fock level, $\Pi(\mathbf{r})$ and $(\rho(\mathbf{r})/2)^2$ has the same value at every point of space and their learning curve should overlap perfectly. As reported in upper panel of Figure 2, however, the absolute error of the two fields differ significantly throughout

the learning. In reality the difficulty of the learning exercise is very similar for the two fields as shown using the relative percentage error. The difference between the two metrics is the direct manifestation of an underlying physical effect: the presence of electron correlation at CASSCF level decreases in average the amplitudes of $\Pi(\mathbf{r})$ with respect to those of two independent particles.

Overall, the performance of the model on the test set is more than satisfying with a mean relative error at the full training set (2550 molecules) of only $3.32 \cdot 10^{-5}$ % for $\Pi(\mathbf{r})$ and $3.14 \cdot 10^{-5}$ % for $(\rho(\mathbf{r})/2)^2$. These errors refer to the deviation from the fitted fields and, thus, represent the machine learning error only. Nevertheless, the fitting error with respect to the *ab-initio* reference (see Section II) is orders of magnitudes lower than the prediction error, which allows the direct comparison between prediction and *ab-initio* results.

$C_0{}^2$: **0.87**
R.E. = 3.15 · 10^{-5} %

Reference   Predicted

$C_0{}^2$: **0.95**
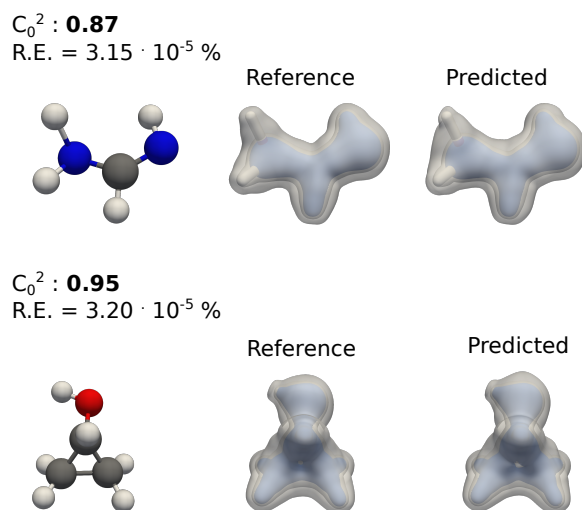R.E. = 3.20 · 10^{-5} %

Reference   Predicted

FIG. 3. Predicted on-top pair density of two illustrative molecules from test set molecules. The reference field (target of the learning) is $\Pi(\mathbf{r})$ fitted on $\phi^{\mathrm{OTPD}}(\mathbf{r})$. Overall, the main real-space features and shape of the on-top pair density is well reproduced by the predictions. Isovalues: 0.01 $e^2 \cdot$ Bohr^{-6} (blue), 0.005 $e^2 \cdot$ Bohr^{-6} (innermost, gray), and 0.001 $e^2 \cdot$ Bohr^{-6} (outermost, gray). The weight of the dominant configuration (C_0^2) and the relative error of the prediction (R.E.) is reported for each molecule.

Quantifying the prediction error through a well-defined numerical metric, such as the absolute and relative error, is an objective procedure to estimate the deviation from the targeted fields, but it tells only little about the impact of such an error on the shape of $\Pi(\mathbf{r})$, its real-space properties and its applicability in methods such as MC-PDFT. For this reason, we report in Figure 3 two illustrative examples of the quality of predicted on-top pair densities in real-space. The two molecules were selected from the test set following two distinct criteria. First, both molecules are characterized by a relative prediction error similar to the mean on the test set (last point of the learning curve). Second, the first molecule is characterized by an evident multiconfigurational character (CASSCF C_0^2 : 0.87), while the electronic structure of the other is already well-described by its most dominant configuration only (C_0^2 : 0.95).

In both molecules, regardless of their multireference character, the main features and the overall shape of the on-top pair density is well reproduced by the predictions. This is especially true for the highest amplitudes, closer to the nuclei (blue parts in Figure 3, while the prediction error is somewhat more evident in the low-$\Pi(\mathbf{r})$ regions (see *e.g.* the primary amine group of the first molecule). The slight degradation of the quality of the regression further from the nuclei has been already observed in the case of the electron density and it has to be attributed in part to the incompleteness of the basis set and in part to fact that the machine learning algorithm tends to minimize the overall error by reducing the discrepancies in the high-amplitude regions. Overall, the quality of the regression is largely sufficient for our goal to apply the predicted on-top pair densities as real-space metrics, as discussed in the following section.

IV. A VISUAL METRIC OF ELECTRON CORRELATION AND BOND BREAKING

The ability to predict from the molecular structure only both the on-top pair densities [$\Pi(\mathbf{r})$] and squared half-densities [$(\rho(\mathbf{r})/2)^2$] at CASSCF level promotes their application as real-space metrics for the evaluation of electron correlation effects and the characterization of bond-types and bond-breaking. This kind of application, similar in spirit to other well-established scalar fields based on the electron density,[56–61] has been proposed and meticulously analyzed for simple molecules at CASSCF level by Carlson, Truhlar and Gagliardi.[25,26] The key quantity to be computed is the on-top ratio, which is defined as:

$$R(\mathbf{r}) = \frac{\Pi(\mathbf{r})}{[\rho(\mathbf{r})/2]^2} \qquad (5)$$

Figure 4 reports the on-top ratio computed from the predicted on-top pair densities and squared half-densities of the same two molecules presented in Figure 3. For visualization purposes the on-top ratio is projected on the surface of squared half-densities for the two molecules.

Although, from the quantitative perspective, it can be noticed that the on-top ratio from the predicted quantities is slightly underestimated on the hydrogen atom of the first molecule, overall it reproduces well both the shape and the qualitative distribution of the reference field. More importantly, because of its correct distribution, the chemical interpretation of the predicted on-top ratio is still possible. As shown in the Figure, the on-top ratio is very close to 1 (HF limit) everywhere in the second molecule, where a single electronic configuration dominates the ground-state. The few regions where the ratio decreases slightly (0.85) are a consequence of dynamic electron correlation, which is partially included along with static correlation in the CASSCF computation. In contrast, the first molecule has a significant multiconfigurational character, originating from the significantly elon-
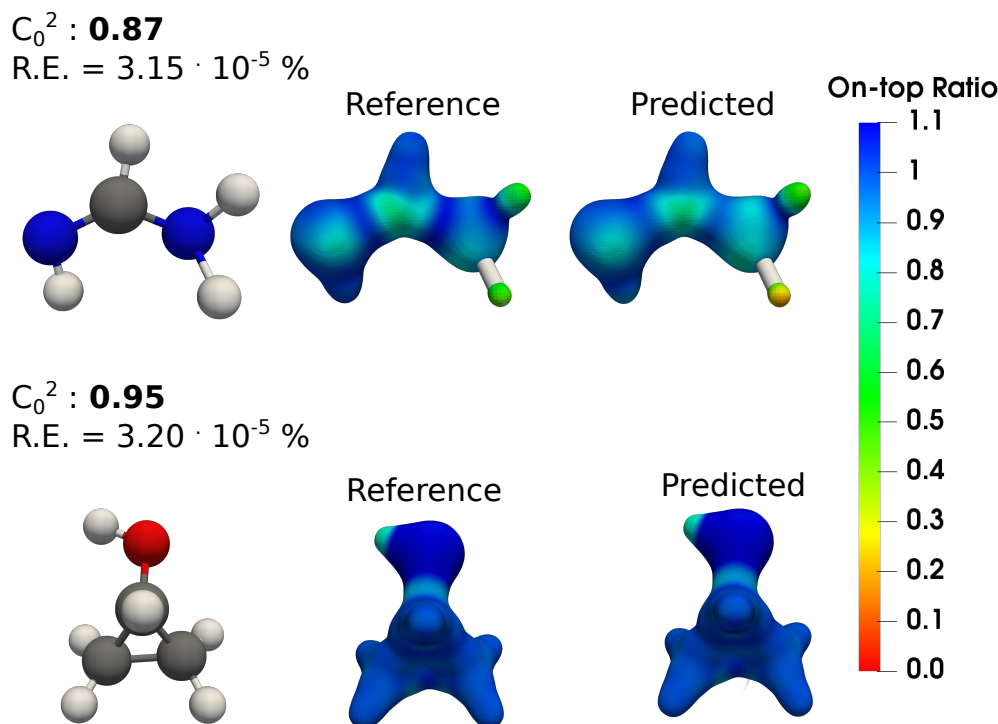
FIG. 4. Predicted on-top ratio [color-code, $R(\mathbf{r})$] for two illustrative molecules of the test set projected on their squared densities (isovalue $0.025\ e^2 \cdot Bohr^{-6}$). The reference R is computed using the fitted on-top pair densities and squared half-densities.

gated N-H bonds in the primary amine (1.24 Å and 1.43 Å respectively, while the standard length is ~ 1.00 Å). In the region of these two hydrogen, the on-top ratio drops significantly and approaches zero. This behavior is not surprising, as at most 1 electron (homolytic bond breaking) would be present on each of the dissociating hydrogens and the $\Pi(\mathbf{r})$ reflects the probability of two electrons sharing the same position in space. These examples concretely demonstrate how the predicted on-top pair densities and squared half-densities are sufficiently accurate to access valuable real-space information about correlation effects and bond-breaking in molecules.

V. CONCLUSION

The centrality of the on-top pair density for the description of electron correlation in atoms, molecules and materials is well reflected by the immense research effort put in the analysis of its properties and applications. In this work, we have proposed a machine learning model of $\Pi(\mathbf{r})$ capable of yielding accurate on-top pair densities given the molecular structure only. Bypassing the demanding CASSCF computations, our model becomes a powerful tool for the visualization and the analysis of electron correlation effects (both static and dynamic). In principles, the predicted densities and on-top pair densities could be also used as ingredients for electronic structure methods such as MC-PDFT. Nevertheless, this kind of application is substantially limited by the necessity to use the

wavefunction to compute the kinetic energy contribution to the total electronic energy.

The regression of $\Pi(\mathbf{r})$ is also the first concrete demonstration of the generalizability to any real-space field of the SA-GPR framework. While the machine learning machinery remains the same regardless of the chosen target, it is imperative to develop efficient decomposition schemes capable of minimizing the fitting error of the target field onto a single basis set. In this respect, we have reported the construction of a specialized basis set for the fitting of the on-top pair density and any general field which behaves as the square of the electron density.

Taking advantage of its locality, the current model is transferable and allows the regression of the on-top pair density of organic molecules of any size, as long as their local atomic environments are contained within the GDB11-AD-3165 database. The construction rules of GDB11-AD-3165 (*i.e.* singlet, neutral molecules only with maximum 6 heavy atoms) result into some limitations for the transferability of the current model, since structural patterns such as fused aromatic rings and electronic patterns such as open-shell radicals and charged species are not included. Transition metal complexes are also not yet compatible with the current form of the framework. Nevertheless, these limitations could be easily lifted in the future by the optimization of specialized basis functions for the transition metal centers and the extension of the training set to any compound of interest, including charged compounds, radicals and transition metal complexes.

As a second advantage of its locality, the model is not restricted to molecules in gas-phase, but it could be applied as a tool to identify and visualize the effects of strong correlation in molecular materials, organic crystals and molecules in solution. This kind of application demands, however, the construction of a training set *ad-hoc*, since it would be crucial for the model to capture the complex effects of intermolecular interactions on $\Pi(\mathbf{r})$.

ACKNOWLEDGMENTS

DATA AVAILABILITY STATEMENT

The data, the basis set exponents and the model that support the findings of this study are openly available in Materials Cloud.

Appendix A: Computational Details

The molecular geometries for all compounds were taken as published in the GDB11-AD-3165 database.[39] The on-top pair densities and the squared half-densities were computed at CASSCF[62]/cc-pVTZ[63] level as implemented in OpenMolcas.[64] The size and the composition of the active spaces were taken from the original reference.[39] The visualization of all the scalar fields and the generation of Figures was performed with a slightly modified version of the Paraview 5.6.0 software.[65] Basis set generation and optimization were performed with an in-house code, which is provided on GitHub (*put here the github*). The projection integrals of the on-top pair density onto the specialized basis set and the basis overlap integrals were computed numerically on a (99, 590) MHL[66]-Lebedev[67] grid with a crowding factor of 5.0.

The tensorial λ-SOAP kernels[41] were computed with the following parameters: environment cutoff $r_{cut} = 4$ Å, Gaussian smearing $\sigma = 0.3$ Å, angular cutoff $l_{cut} = 6$, radial cutoff $n_{cut} = 8$, environmental kernel exponent $\zeta = 2$. A subset of $M = 1000$ reference environments was taken to reduce the dimensionality of the regression problem, and the regularization parameter η was set to 10^{-6}.

REFERENCES

[1] R. McWeeny, *Methods of molecular quantum mechanics* (Academic press, 1992).

[2] J. P. Perdew, A. Savin, and K. Burke, "Escaping the symmetry dilemma through a pair-density interpretation of spin-density functional theory," Phys. Rev. A **51**, 4531–4541 (1995).

[3] P.-O. Löwdin, "Quantum Theory of Many-Particle Systems. I. Physical Interpretations by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configurational Interaction," Phys. Rev. **97**, 1474–1489 (1955).

[4] O. Gunnarsson and B. I. Lundqvist, "Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism," Phys. Rev. B **13**, 4274–4298 (1976).

[5] A. Görling, "Symmetry in density-functional theory," Phys. Rev. A **47**, 2783–2799 (1993).

[6] T. Ziegler, A. Rauk, and E. J. Baerends, "On the calculation of multiplet energies by the hartree-fock-slater method," Theor. Chim. Acta **43**, 261–271 (1977).

[7] J. P. Perdew, M. Ernzerhof, K. Burke, and A. Savin, "On-top pair-density interpretation of spin density functional theory, with applications to magnetism," Int. J. Quantum Chem. **61**, 197–205 (1997).

[8] K. Burke, J. P. Perdew, and M. Ernzerhof, "Why semilocal functionals work: Accuracy of the on-top pair density and importance of system averaging," J. Chem. Phys. **109**, 3760–3771 (1998).

[9] F. Moscardó and E. San-Fabián, "Density-functional formalism and the two-body problem," Phys. Rev. A **44**, 1549–1553 (1991).

[10] F. Moscardó and E. San-Fabián, "A density functional for the correlation energy, deduced in the framework of the correlation factor approach," Int. J. Quantum Chem. **40**, 23–32 (1991).

[11] A. D. Becke, A. Savin, and H. Stoll, "Extension of the local-spin-density exchange-correlation approximation to multiplet states," Theor. Chim. Acta **91**, 147–156 (1995).

[12] B. Miehlich, H. Stoll, and A. Savin, "A correlation-energy density functional for multideterminantal wavefunctions," Mol. Phys. **91**, 527–536 (1997).

[13] J. Gräfenstein and D. Cremer, "The combination of density functional theory with multi-configuration methods – CAS-DFT," Chem. Phys. Lett. **316**, 569–577 (2000).

[14] R. Takeda, S. Yamanaka, and K. Yamaguchi, "Approximate on-top pair density into one-body functions for CAS-DFT," Int. J. Quantum Chem. **96**, 463–473 (2004).

[15] S. Gusarov, P.-Å. Malmqvist *, and R. Lindh, "Using on-top pair density for construction of correlation functionals for multideterminant wave functions," Mol. Phys. **102**, 2207–2216 (2004).

[16] M. Hapka, E. Pastorczak, A. Krzemińska, and K. Pernal, "Long-range-corrected multiconfiguration density functional with the on-top pair density," J. Chem. Phys. **152**, 094102 (2020).

[17] G. Li Manni, R. K. Carlson, S. Luo, D. Ma, J. Olsen, D. G. Truhlar, and L. Gagliardi, "Multiconfiguration Pair-Density Functional Theory," J. Chem. Theory Comput. **10**, 3669–3680 (2014).

[18] L. Gagliardi, D. G. Truhlar, G. Li Manni, R. K. Carlson, C. E. Hoyer, and J. L. Bao, "Multiconfiguration Pair-Density Functional Theory: A New Way To Treat Strongly Correlated Systems," Acc. Chem. Res. **50**, 66–73 (2017).

[19] A. M. Sand, C. E. Hoyer, K. Sharkas, K. M. Kidder, R. Lindh, D. G. Truhlar, and L. Gagliardi, "Analytic Gradients for Complete Active Space Pair-Density Functional Theory," J. Chem. Theory Comput. **14**, 126–138 (2018).

[20] J. L. Bao, L. Gagliardi, and D. G. Truhlar, "Self-Interaction Error in Density Functional Theory: An Appraisal," J. Phys. Chem. Lett. **9**, 2353–2358 (2018).

[21] S. Ghosh, A. L. Sonnenberger, C. E. Hoyer, D. G. Truhlar, and L. Gagliardi, "Multiconfiguration Pair-Density Functional Theory Outperforms Kohn–Sham Density Functional Theory and Multireference Perturbation Theory for Ground-State and Excited-State Charge Transfer," J. Chem. Theory Comput. **11**, 3643–3649 (2015).

[22] R. K. Carlson, G. Li Manni, A. L. Sonnenberger, D. G. Truhlar, and L. Gagliardi, "Multiconfiguration Pair-Density Functional Theory: Barrier Heights and Main Group and Transition Metal Energetics," J. Chem. Theory Comput. **11**, 82–90 (2015).

[23] J. L. Bao, S. O. Odoh, L. Gagliardi, and D. G. Truhlar, "Predicting Bond Dissociation Energies of Transition-Metal Compounds by Multiconfiguration Pair-Density Functional Theory and Second-Order Perturbation Theory Based on Correlated Participating Orbitals and Separated Pairs," J.

Chem. Theory Comput. **13**, 616–626 (2017).

[24]C. E. Hoyer, S. Ghosh, D. G. Truhlar, and L. Gagliardi, "Multiconfiguration Pair-Density Functional Theory Is as Accurate as CASPT2 for Electronic Excitation," J. Phys. Chem. Lett. **7**, 586–591 (2016).

[25]R. K. Carlson, D. G. Truhlar, and L. Gagliardi, "On-Top Ratio for Atoms and Molecules," J. Phys. Chem. A **123**, 8294–8304 (2019).

[26]R. K. Carlson, D. G. Truhlar, and L. Gagliardi, "On-Top Pair Density as a Measure of Left–Right Correlation in Bond Breaking," J. Phys. Chem. A **121**, 5540–5547 (2017).

[27]R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, "Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach," J. Chem. Theory Comput. **11**, 2087–2096 (2015), arXiv:1503.04987.

[28]F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, "Bypassing the Kohn-Sham equations with machine learning," Nat. Comm. **8**, 872 (2017).

[29]A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, "Machine learning unifies the modeling of materials and molecules," Sci. Adv. **3**, e1701816 (2017), arXiv:1706.00179.

[30]S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, "Towards exact molecular dynamics simulations with machine-learned force fields," Nat. Comm. **9**, 3887 (2018).

[31]M. Welborn, L. Cheng, and T. F. Miller, "Transferability in Machine Learning for Electronic Structure via the Molecular Orbital Basis," J. Chem. Theory Comput. **14**, 4772–4779 (2018), arXiv:1806.00133.

[32]J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. E. Roitberg, "Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning," Nat. Commun. **10**, 2903 (2019).

[33]R. Fabregat, A. Fabrizio, B. Meyer, D. Hollas, and C. Corminboeuf, "Hamiltonian-Reservoir Replica Exchange and Machine Learning Potentials for Computational Organic Chemistry," J. Chem. Theory Comput. **16**, 3084–3094 (2020).

[34]Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby, and T. F. Miller, "OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features," J. Chem. Phys. **153**, 124111 (2020), arXiv:2007.08026.

[35]T. Husch, J. Sun, L. Cheng, S. J. R. Lee, and T. F. Miller, "Improved accuracy and transferability of molecular-orbital-based machine learning: Organics, transition-metal complexes, non-covalent interactions, and transition states," (2020), arXiv:2010.03626 [physics.chem-ph].

[36]A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, "Transferable Machine-Learning Model of the Electron Density," ACS Cent. Sci. **5**, 57–64 (2019).

[37]A. Fabrizio, A. Grisafi, B. Meyer, M. Ceriotti, and C. Corminboeuf, "Electron density learning of non-covalent systems," Chem. Sci. **10**, 9424–9432 (2019).

[38]A. Fabrizio, K. Briling, A. Grisafi, and C. Corminboeuf, "Learning (from) the Electron Density: Transferability, Conformational and Chemical Diversity," Chim. Int. J. Chem. **74**, 232–236 (2020).

[39]C. Duan, F. Liu, A. Nandy, and H. J. Kulik, "Data-Driven Approaches Can Overcome the Cost–Accuracy Trade-Off in Multireference Diagnostics," J. Chem. Theory Comput. **16**, 4373–4387 (2020).

[40]F. Weigend, "A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency," Phys. Chem. Chem. Phys. **4**, 4285–4291 (2002).

[41]A. Grisafi, D. M. Wilkins, G. Csányi, and M. Ceriotti, "Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems," Phys. Rev. Lett. **120**, 036002 (2018).

[42]P. D. Walker and P. G. Mezey, "Molecular electron density Lego approach to molecule building," J. Am. Ceram. Soc. **115**, 12423–12430 (1993).

[43]P. D. Walker and P. G. Mezey, "Ab Initio Quality Electron Densities for Proteins: A MEDLA Approach," J. Am. Chem. Soc. **116**, 12022–12032 (1994).

[44]T. E. Exner and P. G. Mezey, "Ab Initio-Quality Electrostatic Potentials for Proteins: An Application of the ADMA Approach," J. Phys. Chem. A **106**, 11791–11800 (2002).

[45]T. E. Exner and P. G. Mezey, "Ab initio quality properties for macromolecules using the ADMA approach," J. Comput. Chem. **24**, 1980–1986 (2003).

[46]Z. Szekeres, T. Exner, and P. G. Mezey, "Fuzzy fragment selection strategies, basis set dependence and HF-DFT comparisons in the applications of the ADMA method of macromolecular quantum chemistry," Int. J. Quantum Chem. **104**, 847–860 (2005).

[47]G. L. Stoychev, A. A. Auer, and F. Neese, "Automatic Generation of Auxiliary Basis Sets," J. Chem. Theory Comput. **13**, 554–562 (2017).

[48]S. Huzinaga and B. Miguel, "A comparison of the geometrical sequence formula and the well-tempered formulas for generating GTO basis orbital exponents," Chem. Phys. Lett. **175**, 289–291 (1990).

[49]F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy." Phys. Chem. Chem. Phys. **7**, 3297–305 (2005).

[50]C. G. Broyden, "The convergence of a class of double-rank minimization algorithms 1. general considerations," IMA J. Appl. Math. **6**, 76–90 (1970).

[51]R. Fletcher, "A new approach to variable metric algorithms," Comput. J. **13**, 317–322 (1970).

[52]D. Goldfarb, "A family of variable-metric methods derived by variational means," Math. Comput. **24**, 23–23 (1970).

[53]D. F. Shanno, "Conditioning of quasi-Newton methods for function minimization," Math. Comput. **24**, 647–647 (1970).

[54]N. C. Handy and A. J. Cohen, "Left-right correlation energy," Mol. Phys. **99**, 403–412 (2001).

[55]D. K. W. Mok, R. Neumann, and N. C. Handy, "Dynamical and Nondynamical Correlation," J. Phys. Chem. **100**, 6225–6230 (1996).

[56]R. F. W. Bader, "A quantum theory of molecular structure and its applications," Chem. Rev. **91**, 893–928 (1991).

[57]R. Bader, *Oxford University Press*, edited by C. F. Matta and R. J. Boyd (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2007).

[58]E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, and W. Yang, "Revealing Noncovalent Interactions," J. Am. Chem. Soc **132**, 6498–6506 (2010).

[59]J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, and W. Yang, "NCIPLOT: A Program for Plotting Noncovalent Interaction Regions," J. Chem. Theory Comput. **7**, 625–632 (2011).

[60]P. de Silva and C. Corminboeuf, "Simultaneous Visualization of Covalent and Noncovalent Interactions Using Regions of Density Overlap," J. Chem. Theory Comput. **10**, 3745–3756 (2014).

[61]S. Grimme and A. Hansen, "A Practicable Real-Space Measure and Visualization of Static Electron-Correlation Effects," Angew. Chemie Int. Ed. **54**, 12308–12313 (2015).

[62]B. O. Roos, P. R. Taylor, and P. E. Sigbahn, "A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach," Chem. Phys. **48**, 157–173 (1980).

[63]T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," J. Chem. Phys. **90**, 1007–1023 (1989).

[64]I. Fdez. Galván, M. Vacher, A. Alavi, C. Angeli, F. Aquilante, J. Autschbach, J. J. Bao, S. I. Bokarev, N. A. Bogdanov, R. K. Carlson, L. F. Chibotaru, J. Creutzberg, N. Dattani, M. G. Delcey, S. S. Dong, A. Dreuw, L. Freitag, L. M. Frutos, L. Gagliardi, F. Gendron, A. Giussani, L. González, G. Grell, M. Guo, C. E. Hoyer, M. Johansson, S. Keller, S. Knecht, G. Kovačević, E. Källman, G. Li Manni, M. Lundberg, Y. Ma, S. Mai, J. P. Malhado, P. Å. Malmqvist, P. Marquetand, S. A. Mewes, J. Norell, M. Olivucci, M. Oppel, Q. M. Phung, K. Pierloot, F. Plasser, M. Reiher, A. M. Sand, I. Schapiro, P. Sharma, C. J. Stein, L. K. Sørensen, D. G. Truhlar, M. Ugandi, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, O. Weser, T. A. Wesołowski, P.-O. Widmark, S. Wouters, A. Zech, J. P. Zobel, and R. Lindh, "OpenMolcas: From Source Code to Insight," J. Chem. Theory Comput. **15**, 5925–5964 (2019).

[65]J. Ahrens, B. Geveci, and C. Law, "ParaView: An End-User Tool for Large-Data Visualization," in *Vis. Handb.* (Elsevier, 2005) pp. 717–731.

[66]C. W. Murray, N. C. Handy, and G. J. Laming, "Quadrature schemes for integrals of density functional theory," Mol. Phys. **78**, 997–1014 (1993), 00268979300100651.

[67]V. I. Lebedev, "Quadrature formulas of orders 41, 47, and 53 for the sphere," Russ. Acad. Sci. Dokl. Math. **45**, 587–592 (1992).