We present an interpretation of Fermi-orbital descriptors (FODs) and argue that these descriptors carry chemical bonding information. We show that a bond order derived from these FODs agrees well with reference values, and highlight that optimized FOD positions used within the Fermi-Löwdin orbital self-interaction correction (FLO-SIC) method correspond to expectations from Linnett’s double-quartet theory, which is an extension of Lewis theory. This observation is independent of the underlying exchange-correlation functional, which is shown using the local spin density approximation, the Perdew–Burke–Ernzerhof generalized gradient approximation (GGA), and the strongly constrained and appropriately normed meta-GGA. To make FOD positions generally accessible, we propose and discuss four independent methods for the generation of Fermi-orbital descriptors, their implementation as well as their advantages and drawbacks. In particular, we introduce a re-implementation of the electron force field, an approach based on the centers of mass of orbital densities, a Monte Carlo-based algorithm, and a method based on Lewis-like bonding information. All results are summarized with respect to future developments of FLO-SIC and related methods. © 2019 The Authors. Journal of Computational Chemistry published by Wiley Periodicals, Inc.

DOI: 10.1002/jcc.26062

Introduction

Density functional theory (DFT)\(^1,2\) is one of the most commonly used electronic structure methods. Its popularity comes from its efficiency to calculate properties with reasonable accuracy for a range of systems including atoms, molecules,\(^3–6\) clusters,\(^7\) and condensed matter.\(^8–15\) While DFT is in principle exact, the exchange-correlation functional \(E_{\text{xc}}\) needs to be approximated to solve the Kohn–Sham (KS) equations in practical calculations. These density functional approximations (DFAs) often work well, but are known to produce incorrect results for rather simple cases like the dissociation of the \(\text{H}_2\) molecule\(^16\) or band gaps in semiconductors that are far from experimental values.\(^17\)

The exchange hole\(^18\) (also referred to as the Fermi hole) and the correlation hole (also referred to as the Coulomb hole) are important properties in electronic structure theories, which are often used to discuss exchange-correlation effects.\(^19–24\) The exchange hole\(^25\) determines the decrease in the probability of finding an electron at point \(\mathbf{r}\), given an electron of the same spin at \(\mathbf{r}'\). The exchange hole is negative definite everywhere and integrates to \(-1\). In contrast, the correlation hole integrates to \(0\). The correlation hole corresponds to the decrease in the probability of finding an electron at point \(\mathbf{r}\), given an electron at \(\mathbf{r}'\), due to their charge repulsion. These are constraints an exact DFA should obey.

These holes have been extensively discussed within wave function theories, for example, Hartree–Fock (HF)\(^26,27\) theory.\(^25,28–30\) There, a separation of the exchange and correlation parts of the energy is possible so that the Fermi and Coulomb hole can be unambiguously defined. Please note that exchange-correlation functionals treat exchange and correlation effects together. Thus, one obtains a density including both exchange and correlation effects and an exchange-correlation hole where the separation into exchange and correlation holes is not straightforward. To a large extent, the mentioned failures of DFAs can be linked to the so-called self-interaction error (SIE), describing the interaction of an electron with itself. This error occurs due to an incomplete cancellation of the Coulomb self-repulsion by all local or semi-local \(E_{\text{xc}}\) like the local (spin) density approximation (LSDA), generalized gradient approximations (GGAs) or meta-GGAs. To remove this error, Perdew and Zunger (PZ)\(^31\) proposed a self-interaction correction (SIC) scheme,

\[
E_{\text{PZ-SIC}} = E_{\text{KS-DFT}} - E_{\text{SIC}},
\]

with \(E_{\text{KS-DFT}}\) being the KS-DFT energy functional and \(E_{\text{SIC}}\) being the self-interaction correction functional given by

---

\(1\) S. Schwalbe, L. Fiedler, J. Kraus, T. Hahn, J. Kortus
TU Freiberg, Institute of Theoretical Physics, Leipziger Str. 23, D-09599, Freiberg, Germany
E-mail: schwalbe@physik.tu-freiberg.de

\(2\) K. Trepte, A. I. Johnson, J. E. Peralta, K. A. Jackson
Department of Physics, Central Michigan University, Mount Pleasant, Michigan, 48859

\(3\) L. Fiedler
Freiberg Instruments GmbH, Delfter Str. 6, D-09599, Freiberg, Germany
Contract Grant sponsor: U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences; Contract Grant number: #DE-SC0018331;
Contract Grant sponsor: Deutsche Forschungsgemeinschaft; Contract Grant number: KO 1924/9-1

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Journal of Computational Chemistry published by Wiley Periodicals, Inc.
\[ E_{\text{FLO-SIC}} = \sum_{\sigma} \left( E_{\text{H}}[n_{\sigma}^i] + E_{\text{XC}}[n_{\sigma}^i, 0] \right). \] (2)

Here, \( E_{\text{H}}[n_{\sigma}^i] \) is the underlying exchange-correlation functional evaluated for the one-electron density \( n_{\sigma}^i \) and \( E_{\text{XC}}[n_{\sigma}^i, 0] \) is the corresponding Coulomb term. Considering the one-electron density limit, SIC calculations reproduce HF results, which are exact for this limit.\(^{31,32}\) For many-electron systems, the application of the PZ scheme in any self-consistent form offered significant improvements with respect to the underlying density functional, for example, for ionization potentials \(^{33-36}\) and (hyper-)polarizabilities.\(^{37-38}\)

It has been shown that localized orbitals deliver lower self-interaction corrected total energies and, accordingly, play an essential role in a proper self-interaction correction in DFT.\(^{31,39,40}\) The advantages and drawbacks of each method will be discussed in detail—an approach using an electron force field, a quantum mechanical approach based on the centroids of localized one-body densities, a Monte Carlo-based algorithm, and an approach using information from Lewis structural formulae. The advantages and drawbacks of each method will be discussed with respect to automation, speed, scope, extensibility, and insensitivity. These criteria will be introduced in detail later.

A short introduction to the concept of bond orders and to the nearly forgotten Linnett double-quartet (LDQ) theory\(^{59}\) is given in the next section.

### Relation Between FODs and Chemical Bonding Information

The adiabatic connection formula \(^{60}\) relates the exchange-correlation energy to a coupling-constant integrated exchange-correlation hole. The exchange-correlation hole visualizes electron correlation in an inhomogeneous material by measuring the change in density from its mean value at each point in a system with respect to a particular reference point. The exchange-correlation energy per particle can be interpreted as the interaction energy of the particle with its exchange-correlation hole. The exchange hole refers to a noninteracting ground state and contains only correlation due to the Pauli principle. It is often the dominant part of the total exchange-correlation hole.\(^{61}\) In the case of covalent bonds, the exchange hole is not very dependent on the reference position and can be approximated by localized orbitals near the bond center.\(^{61}\) This connection between chemical bonding and the exchange-correlation hole has also been discussed by Giesbertz et al.\(^{62}\) The FODs are used to construct the Fermi orbitals, whose squares can be used to approximate the exchange-correlation hole.\(^{19,29,41}\) This indicates that the optimized positions of the FODs in FLO-SIC may carry bonding information. In this section, we set out to investigate this in more detail.

In the chemistry community, it is common practice to use the concept of bond order.\(^{63,64}\) Localized orbitals (e.g., NBO\(^{65}\)) and other concepts\(^{66,67}\) can be used to determine bond orders in a post-processing fashion. In FLO-SIC, one can define a simple bond order (BO\(_{\text{FOD}}\)) as

\[ \text{BO}_{\text{FOD}} = \frac{m_{\text{electrons}}}{n_{\text{centers}}}, \] (4)

where \( m \) is the number of electrons involved in the bond and \( n \) is the number of atoms that belong to this bond. In this
context, we attribute one FOD to one electron. The BO\textsubscript{FOD} is based on counting the FODs between bonded nuclear centers. A detailed example of the FOD bond order for \textit{N}_2 is given in Figure 1. A BO\textsubscript{FOD} of 1 represents a single bond (e.g., C=C), 2 represents a double bond (e.g., C≡C) and 3 represents a triple bond (C≡C≡C). All of the mentioned bond orders are represented well by Lewis theory, the conventional framework for describing bonding in chemistry that is based on the octet rule and popularly visualized in Lewis dot diagrams. However, there are other bond orders that are not properly described by Lewis theory. For example, 2.5 represents a two-center, three-electron bond and 1.5 represents a two-center, three-electron bond. In addition, we distinguish between the commonly known double bond (e.g., C=C) consisting of two electrons in each spin channel and the uncommon double bond consisting of three electrons of one spin channel and one electron of the other spin channel. We have chosen a shorthand notation in which two-electron bonds are represented by solid lines, whereas one-electron bonds are represented by dotted lines.

A systematic analysis of recently published FLO-SIC results for the G2-1 benchmark set [36] allows us to determine the FOD bond order for all molecules of this set. The results are summarized in the Supporting Information. FLO-SIC provides essentially the same bond order as is provided by Lewis structures for systems with straightforward bonding (51 molecules in total). For four systems (see Table 1), literature references give conflicting bond order values.

All of these systems can be described as radicals, that is, with a structure containing at least one unpaired electron. In such cases, the Lewis structure is not clear, and different sources (e.g., CCCBDB, PubChem, ChemSpider) give different structures and hence different bond orders. In view of this, it is unclear which reference value should be taken for comparison (see Supporting Information for a comparison between different literature values). There might be a need for more consistent bond order reference values, but this is beyond the scope of this work.

While FLO-SIC predicts, in agreement with Lewis theory, a bond order of two for the O\textsubscript{2} molecule (see Table 1), it is obvious that the FLO-SIC electronic geometry is not what would be expected from the conventional Lewis structure. FLO-SIC predicts a double bond consisting of three electrons of one spin channel and one electron of the other spin channel. We use this example to recall a nearly forgotten theory named LDQ theory [56–59]. Whereas Lewis never included the spin in his famous model, it plays a critical role in Linnett theory. For Linnett, the octet consists of two quartets, each only containing electrons of the same spin, giving the theory its name. LDQ does not contradict Lewis theory, but rather delivers additional insights where Lewis theory fails.

According to Linnett, the spatial distribution of electrons is determined by two different correlation effects. [57,58,71]. One of them is the Pauli exclusion principle, which leads to the tendency of electrons having parallel spins to keep apart, and favors the spatial proximity of electrons with antiparallel spins. [57,58,71]. The second correlation effect is that all electrons carry the same negative charge and therefore aim to keep apart from each other because of Coulomb repulsion. Consequently, four electrons of the same spin experience negative charge correlation on top of negative spin correlation. These electrons try to arrange themselves in a way that maximizes their distance from each other. [57,58,71]. Geometrically, this is achieved by occupying the corners of a regular tetrahedron (4 sp\textsuperscript{3} orbitals). [57,58,71]. The overall correlation between the two spin quartets is assumed to be small, as the positive spin correlation partly cancels the negative charge correlation. Therefore, the quartets and their tetrahedra keep a certain degree of independence from each other concerning their orientations around the nucleus. [57,58,71]. Linnett implicitly argues that a relatively small deviation from the strictly regular tetrahedra might be energetically favorable in certain cases. [58, 58]

In general, LDQ structures are constructed by arranging the electron tetrahedra in a way that minimizes inter-electronic repulsion by spin and charge, while still maintaining the octet rule. [57,58,71]. Linnett structures can be written similarly to their Lewis counterparts. In contrast to Lewis, however, each simplified Linnett structure assigns every single electron to either a dot or a cross, depending on its spin. [57,58,71]. Linnett and Lewis structures are expected to be the same in straightforward cases like CH\textsubscript{4}, [59] where the bonding arrangement dictates that the up and down spin quartets should be arranged in the same way. Various details and additional aspects are discussed in the bachelor thesis of Kraus. [64].

---

Figure 1. FOD bond order BO\textsubscript{FOD}. The N\textsubscript{2} molecule is used as an example consisting of the nuclear centers (A, B) with their respective FODs (shown as small spheres). Region a describes where lone pair FODs are located. These FODs are associated with one center, but are not spatially close to the nucleus. Region b describes the core FODs. Core FODs are spatially very close to one center, in contrast to valence FODs. Here, spatially close means that the distance of the core FODs is much smaller than the covalent radius of the corresponding atom. Region c describes the bond region, which contains the valence FODs of connected centers (in this case A and B). Counting all m FODs (up, down) in region c for a specific n center arrangement gives the FOD bond order (see eq. (4)). For the N\textsubscript{2} example, we count two centers (n = 2) as well as three up FODs and three down FODs in region c (m = 6). The resulting FOD bond order is m/n = 6/2 = 3, which is interpreted as a triple bond. The three bond FODs form an equilateral triangle around the bond axis. All lone pair and bond FODs form the set of valence FODs (color code: nuclei [blue], up FODs [green], down FODs [red]). Note: There are cases where up and down FODs are located at the same position; thus, the resulting FODs are overlapping and only one color is seen. In the lower part of the figure, the atoms have been omitted to show the entire electronic geometry, including the core FODs. [Color figure can be viewed at wileyonlinelibrary.com]
In the following, we will call attention to analogies between Linnett’s quasi-classical electron positions and the arrangements of optimized FOD positions. Our discussion will be qualitative in nature, with a focus on general structural motifs. Linnett proposed a LDQ structure for the O$_2$ molecule that is in excellent agreement with the FLO-SIC electronic geometry (see Supporting Information). For the CO$_2$ molecule, Linnett compared six different structures [56,58] (including the Lewis structure). The structure that he predicted to be most stable was not the Lewis structure. In fact, it was a structure which (similarly to Linnett’s O$_2$ structure) does not show the common double bonds (i.e., C=O, two up and two down FODs). Instead, it features two uncommon double bonds (i.e., three up FODs and one down FOD or vice versa). For both structures, the bond order is identical and charge neutrality is ensured, but for the Linnett structure, the quasi-electron positions are further apart from one another.

We found that Linnett’s proposal is in agreement with FLO-SIC results (see Fig. 2). The Linnett-like FOD arrangement for CO$_2$ is energetically preferred by about 2m$E_h$ ($1E_h = 27.2112$ eV) over the Lewis structure within FLO-SIC LDA (see Supporting Information as well). This is not an isolated occurrence. We found that bonding situations as described by the electronic geometry in FLO-SIC generally agree with LDQ theory, especially in cases where the predictions differ from Lewis theory.

**Fermi-Orbital Descriptor Generators**

As shown above, FODs carry bonding information, and they are needed to carry out FLO-SIC calculations, thus it is important to have suitable methods to generate them. Fully optimizing the FODs is numerically demanding, as the evaluation of FOD forces, which are defined as the derivatives of the total energy with respect to FOD positions, is time-consuming. The numerical effort increases with the number of electrons, that is, the number of FODs. Therefore, it is desirable to have initial FODs that are close to the optimal positions to speed up FLO-SIC calculations.

### Table 1

| System | n-c-m-e | FLO-SIC LDA | FLO-SIC PBE | REF |
|--------|---------|-------------|-------------|-----|
| HCO    | n-c-m-e | 1/2-2.5     | 1/2-2.5     | 1/2-2.5 |
| O2     | n-c-m-e | 1/2-2.5     | 1/2-2.5     | 1/2-2.5 |
| SiC    | n-c-m-e | 1/2-2.5     | 1/2-2.5     | 1/2-2.5 |

**Figure 2.** Lewis and Linnett structures for the CO$_2$ molecule. The corresponding FOD geometries are shown as well [color code: C (brown), O (red), up FODs (green), down FODs (red)]. Up and down FODs in the Lewis structure are at the same positions, thus only one color (green) is shown. [Color figure can be viewed at wileyonlinelibrary.com]
In the following, four different approaches to achieve this goal will be presented. The first one is based on a classical force field formulation (PyEFF), the second one uses the centers of mass of single-electron densities (PyCOM), the third one is a Monte-Carlo method minimizing the Coulomb repulsion (fodMC), and the final one works with fixed structural motifs derived from Lewis structures (PyLEWIS). To judge the quality of these methods, FODs generated using them are compared to fully optimized FODs for two benchmark test sets.

Finally, we analyze the performance of the four methods with respect to required user input (automation), range of systems the method can be applied to (scope), computational time (speed), additional utility beyond FOD generation (extensibility), and how much the generated FODs depend on the input (insensitivity).

**Python-based electron force field**

One possibility for a fast automatic generation of FODs would be to use force fields, treating the electrons as classical particles. Optimizing the positions of these classical particles is then analogous to a molecular geometry optimization. It could be thought of as classical electronic positions.

Another option is the so-called electron force field (eFF)\(^{[75–78]}\). The eFF computes the total energy and the forces on the electrons and nuclei of the system. Here, the nuclei are represented as point charges and the electrons as spherical Gaussians\(^{[79]}\)

\[
\psi_i(\mathbf{r}, s_i) = \exp\left(-\frac{(r - x_i)^2}{s_i^2}\right),
\]

that is, as a function of nuclear \(\mathbf{r}\) and electron \(\mathbf{x}\), positions and electron sizes \(s_i\). The electron positions and sizes are varied as parametric functions of the nuclear positions to minimize the total energy of the system given by

\[
E = E_{\text{ke}} + E_{\text{nuc,nuc}} + E_{\text{nuc,elec}} + E_{\text{elec,elec}} + E_{\text{Pauli}}.
\]

This variation is done using a gradient-based algorithm. The terms in eq. (6) describe the sum of kinetic energy \(E_{\text{ke}}\), the Hartree electrostatic energies \(E_{\text{nuc,nuc}}, E_{\text{nuc,elec}}, E_{\text{elec,elec}}\) as well as an antisymmetrization (Pauli) correction \(E_{\text{Pauli}}\). Further details are given in the Supporting Information.

In the eFF, each electron carries a spin (+1/2 or −1/2). For instance, one can calculate a C atom with \(S = 1\) (four up, two down electrons) and \(S = 2\) (five up, one down electrons) in the same way one would handle it within FLO-SIC.

The first formulation of the electron force field\(^{[75–77]}\) (eFF1) may be called a quasi all-electron approximation, because all electrons (core + valence) are treated explicitly. The simple Gaussians used to represent electrons resemble \(s\)-orbitals. Therefore, it is not surprising that the force field becomes less accurate for atoms containing states with higher angular quantum numbers. In the case of eFF1, only elements with \(Z = 1–6\) are described well. The eFF2\(^{[77]}\) formulation tries to overcome this limitation, and has been tested for systems containing elements with \(Z = 1–10\).

The standard eFF1 formulation is available in the eff-code\(^{[75]}\) and in LAMMPS\(^{[80]}\) but the eFF2 formulation is not available yet. The eFF1 formulation is already available in our PyEFF code. As one can see in the Supporting Information, our implementation of eFF1 within PyEFF is in excellent agreement with the reference implementation. PyEFF is written in python and uses standard scientific python packages (e.g., numpy, scipy, sympy, ast). It offers several input and output features to generate FOD input files for the NRLMOL code\(^{[81–87]}\) or our PyFLOSiC implementation.\(^{[88]}\) A typical input for PyEFF is the nuclear geometry, the spin state, and starting values for the electron sizes \(s_i\). Thus, we do not directly use bonding information of the system to generate FOD starting guesses.

Besides the possibility of calculating different spin states, this force field formulation delivers a total energy expression and can therefore be used to describe bond dissociation. Furthermore, PyEFF can be used to calculate ionization potentials and electron affinities. We calculated the \(H_2\) dissociation curve with PyEFF and compared the results to FLO-SIC LDA (see Supporting Information). It is known that DFT is not able to reproduce the correct dissociation limit in contrast to FLO-SIC. The PyEFF bond dissociation curve shows a qualitatively correct description of bond breaking in the stretched-bond region compared to FLO-SIC LDA. However, the energy in the current formulation of the eFF is not sufficient to reach FLO-SIC LDA accuracy. It is interesting to consider how physical the results are, in view of the fact that the optimal classical electron positions suggested by PyEFF are used as guesses for FOD positions, which themselves can be loosely thought of as classical electronic positions.

**Python-based center of mass**

The main assumption for the python-based center of mass (PyCOM) method is that the center of mass (or center of gravity or centroid) of one-particle densities can give reasonable starting values for the FODs. The first step is a standard DFT calculation for a given nuclear geometry, yielding a set of KS orbitals. This requires choosing the exchange-correlation functional, basis set and numerical grid. In PyCOM, the user can choose between using Foster–Boys, Edmiston–Ruedenberg, or Pipek–Mezey localization procedures to transform the KS orbitals (each spin channel separately). The resulting initial FOD guess depends on the choice of the localization method. Lehtola and Jonsson\(^{[89]}\) provide a detailed overview of the different procedures. For our implementation, we used the PySCF\(^{[90]}\) code. The localization procedure in PySCF uses a fast and efficient algorithm based on a co-iterative augmented Hessian method\(^{[91]}\).

After the localization, the orbital densities are calculated on a numerical grid. Finally, the center of mass is calculated and collected for each orbital density. The calculated positions are saved as an xyz file including both nuclei and descriptor positions. The resulting descriptors depend on the exchange-correlation functional, the basis set, the numerical grid, and the localization method. This method scales like a standard density functional
theory calculation, adding the cost of applying a localization method to each spin channel.

This procedure is thought of as a stand-alone post-processing procedure. Having access to orbitals in standard cube-files enables the possibility to use orbitals from any other code. For future use in PyFLOSIC, the method can be sped-up by calculating the trace of the density times the first moment matrix directly using PySCF routines, instead of writing the orbitals on a numerical mesh.

Of the FOD generation methods discussed here, PyCOM is the one with the highest numerical effort (a comparison of timings between PyCOM and fodMC is given in the Supporting Information), but it is applicable to any system that can be studied within density functional theory. The only system-dependent parameters are the atomic coordinates and the spin state.

**Fermi-orbital descriptor Monte Carlo**

This method is based on the fact that FODs typically arrange in structural motifs for atoms and molecules, organized according to atomic subshells (e.g., FODs corresponding to 2s and 2p

| System | \(E_{\text{OPT} - \text{LDA}}\) | \(E_{\text{OPT} - \text{FLO-SIC}}\) | \(E_{\text{GUESS} - \text{FLO-SIC}}\) | \(\Delta E_{\text{FLO-SIC}}\) | \(F_{\text{OPT} - \text{max}}\) |
|--------|----------------|----------------|----------------|----------------|----------------|
| H      | 0.4787         | 0.4897         | 0.4897         | 0.0            | 0.0            |
| He     | 2.8344         | 2.8879         | 2.8917         | 0.0            | 0.0            |
| Li     | 8.3432         | 7.5088         | 7.5091         | 0.3            | 2.1            |
| Be     | 14.4401        | 14.7056        | 14.7066        | 1.0            | 3.7            |
| B      | 24.3534        | 24.7262        | 24.7264        | 0.2            | 0.8            |
| N      | 37.4628        | 37.9554        | 37.9558        | 0.4            | 0.7            |
| C      | 54.1341        | 54.7393        | 54.7395        | 0.2            | 3.2            |
| F      | 99.1107        | 100.0134       | 100.0139       | 0.5            | 0.8            |
| BH₃    | 26.2746        | 26.7271        | 26.7271        | 0.0            | 3.8            |
| C₂H₂   | 76.6249        | 77.6052        | 77.6060        | 0.8            | 4.6            |
| C₂H₄   | 77.8571        | 78.9067        | 78.9077        | 1.0            | 1.2            |
| C₂H₆   | 79.0717        | 80.1959        | 80.1966        | 0.7            | 2.7            |
| C₃H₄   | 116.1748       | 117.7131       | 117.7165       | 3.4            | 2.7            |
| C₃H₆ (Lewis) | 230.1837 | 233.1405        | 233.1432       | 2.7            | 4.1            |
| C₃H₆ (Linnett) | 230.1837 | 233.1428        | 233.1457       | 2.9            | 3.6            |
| CH₂    | 39.4439        | 39.9938        | 39.9938        | 2.4            | 3.2            |
| CH₄    | 40.1186        | 40.6999        | 40.7001        | 0.2            | 0.8            |
| CN     | 91.9501        | 92.9694        | 92.9738        | 4.4            | 3.7            |
| F₂     | 198.3438       | 200.0669       | 200.0896       | 22.7           | 4.2            |
| H₂O₂   | 150.5419       | 152.0218       | 152.0355       | 13.7           | 1.6            |
| HCN    | 92.6541        | 93.7218        | 93.7231        | 1.3            | 2.8            |
| HF     | 99.8476        | 100.7413       | 100.7429       | 1.4            | 3.5            |
| N₂     | 108.6923       | 109.8506       | 109.8558       | 5.2            | 2.5            |
| N₂H    | 163.5555       | 165.2621       | 165.2769       | 14.8           | 1.7            |
| CH₂OH  | 114.8423       | 116.1433       | 116.1474       | 4.1            | 4.4            |
| C₂H₂OH | 153.8044       | 155.6362       | 155.6409       | 4.7            | 1.5            |
| CH₂F   | 138.7822       | 140.2117       | 140.2165       | 4.8            | 1.8            |
| CH₂F₂  | 237.4671       | 239.7362       | 239.7461       | 9.9            | 3.7            |
| CH₂F₃  | 336.1650       | 339.2741       | 339.2884       | 14.3           | 3.4            |
| CF₄    | 434.8488       | 438.7827       | 438.8025       | 19.8           | 4.7            |
| CH₂F   | 138.1115       | 139.5057       | 139.5145       | 8.8            | 1.7            |
| CH₂NH  | 93.8113        | 94.9637        | 94.9667        | 3.0            | 2.2            |
| C₂H₂O  | 152.6040       | 154.3795       | 154.3843       | 4.8            | 2.9            |
| CH₂O   | 113.6201       | 114.8723       | 114.8760       | 3.7            | 3.7            |
| CH₂NO  | 168.6017       | 170.4517       | 170.4627       | 11.0           | 2.8            |
| CH₂SH  | 436.9722       | 439.7369       | 439.7410       | 4.1            | 3.3            |
| C₂H₃SH | 475.9297       | 479.2260       | 479.2306       | 4.6            | 2.5            |
Spherical coordinates \((r, \theta, \phi)\) are used, and all \(N\) positions are initialized at \((r, 0, 0)\). This was done to avoid an initialization near a local minimum. After the initialization, two random numbers \(\alpha\) and \(\beta\) are generated in the interval of \([0,1]\) for all \(N\) points. These serve in the construction of a new position of point \(i (i \in N)\) at step \(n\) using a step size \(a_{\text{step}} = 0.001\).

\[
\theta_i^n = \theta_i^{n-1} + (2\alpha - 1) \pi a_{\text{step}} \\
\phi_i^n = \phi_i^{n-1} + (2\beta - 1) 2\pi a_{\text{step}} \\
x_i^n = r \sin(\theta_i^n) \cos(\phi_i^n) \\
y_i^n = r \sin(\theta_i^n) \sin(\phi_i^n) \\
z_i^n = r \cos(\theta_i^n). 
\]

After all positions have been moved, the Coulomb repulsion is evaluated according to eqs. (7) and (8). If \(r_i < r_i^{n-1}\), the new positions are taken into the next MC step. If not, the old positions are used again. As a note, \(r_i^n\) is set to 10\(^6\). Proof-of-concept calculations for a single sphere were performed, distributing up to 1000 points. The results can be found in the Supporting Information.

### Table 3. FOD benchmark results for PBE. Total energies in \(E_n, \Delta E_{\text{FLO-SIC}} = E_{\text{FLO-SIC}} - E_{\text{FLO-SC}}\) in m\(E_h\), and the absolute maximum FOD force component \(|F_{\text{max}}|\) in \(10^{-14} E_h/\text{bohr}\) using fodMC starting guesses.

| System | \(E_{\text{OPT-PBE}}\) | \(E_{\text{FLO-SIC}}\) | \(\Delta E_{\text{FLO-SIC}}\) | \(F_{\text{max}}\) |
|--------|-----------------|-----------------|-----------------|-----------------|
| H      | -0.4999         | -0.4998         | 0.0             | 0.0             |
| He     | -2.8929         | -2.8926         | 0.0             | 0.0             |
| Li     | -7.4620         | -7.4699         | 0.7             | 3.5             |
| Be     | -14.6296        | -14.6455        | 0.6             | 0.5             |
| B      | -24.6119        | -24.6350        | 0.1             | 0.7             |
| C      | -37.7983        | -37.8094        | 5.4             | 1.0             |
| N      | -54.5349        | -54.5283        | 0.0             | 0.6             |
| F      | -99.6729        | -99.6331        | 0.3             | 0.9             |
| BH₂    | -26.5552        | -26.5837        | 0.0             | 3.1             |
| CH₂    | -77.2560        | -77.2405        | 0.9             | 1.1             |
| CH₄    | -78.5061        | -78.5128        | 0.9             | 2.2             |
| CH₅    | -79.3735        | -79.7596        | 0.7             | 2.3             |
| CH₆    | -117.1376       | -117.1378       | 2.9             | 3.1             |
| CH₇    | -232.0317       | -232.0080       | 2.0             | 2.2             |
| CH₈    | -232.0317       | -232.0109       | 2.0             | 2.9             |
| CH₉    | -39.7916        | -39.7919        | 2.0             | 2.4             |
| CH₁₀   | -40.4669        | -40.4767        | 0.3             | 0.6             |
| CN     | -92.6548        | -92.5836        | 4.0             | 3.0             |
| F₂     | -199.4299       | -199.2786       | 2.0             | 1.0             |
| H₂O₂   | -151.4722       | -151.3725       | 13.6            | 2.5             |
| HCN    | -93.3514        | -93.3109        | 13.6            | 2.5             |
| HF     | -100.3998       | -100.3237       | 1.5             | 0.6             |
| N₂     | -109.4556       | -109.3940       | 4.9             | 4.7             |
| N₂H    | -164.6908       | -164.5421       | 16.3            | 3.1             |
| CH₂OH  | -115.6379       | -115.6012       | 4.3             | 1.0             |
| CH₂OH  | -154.9161       | -154.8901       | 4.8             | 4.6             |
| CH₂F   | -139.6541       | -139.5947       | 4.2             | 2.4             |
| CH₂F₂  | -238.8612       | -238.7263       | 8.5             | 2.9             |
| CH₂F₃  | -338.0780       | -337.8689       | 12.4            | 4.6             |
| CF₄    | -437.2850       | -436.9914       | 17.1            | 3.6             |
| CH₃F   | -138.9183       | -138.9124       | 8.5             | 2.3             |
| CH₃NH  | -94.5245        | -94.5121        | 3.0             | 4.3             |
| C₂H₆O  | -153.7033       | -153.6628       | 3.9             | 2.4             |
| C₂H₅O  | -114.4034       | -114.3617       | 2.6             | 3.0             |
| C₂H₅NO | -169.7550       | -169.6697       | 11.4            | 2.5             |
| C₂H₅SH | -438.5060       | -438.3275       | 7.0             | 3.2             |
| C₂H₅SH | -477.7797       | -477.6126       | 7.4             | 3.0             |
One can imagine such a sphere as a combined orbit for a given set of electrons (e.g., a shell in an atom). With that, this basic MC procedure is extended for the determination of initial FODs. While FODs for atoms are initialized as outlined above, finding FOD positions in molecules is done differently. To be specific, core FODs are optimized using the mentioned MC procedure. A modified ansatz is employed for bond and lone pair FODs, as explained below. The entire approach has been implemented in a FORTRAN routine, called fodMC. Figure 3 shows a simplified outline of the fodMC.

There is a differentiation between spin up (up) and spin down (down) FODs. This allows the treatment of spin-polarized systems and gives a more reasonable picture of the FODs in space. For atoms, the up and down FODs are rotated against each other (after being individually optimized) to formally decrease their Coulomb repulsion (rotations are again based on MC). This leads to inverted tetrahedra of up and down FODs in, for example, Ne (see Supporting Information).

To obtain a good set of starting FODs for molecules, recognizing the presence of bonds and lone pairs is essential. To form bonds, the center between atoms is taken as a starting point. The bond FODs are placed either at this center (single bond) or are initialized perpendicular to the bond axis (multiple bonds). For the latter, MC steps are carried out to equidistantly distribute the bond FODs perpendicular to the bond axis (via rotations). This is analogous to the distribution of the Fermi hole described by Luken. [30]

If the numbers of up and down FODs in the bond are equal, they will be paired. By specifying the number of up and down FODs to be different, one can form, for example, two center—five electron bonds. Furthermore, previous FLO-SIC LDA calculations have shown that FODs always lie closer to the H atom in X–H bonds. Accordingly, the FODs in such bonds are not placed at the midpoint between the atoms, but at 85% of the bond distance toward the H.

For unpaired (e.g., in radicals) and lone pair FODs, a similar ansatz is used. First, all vectors describing bonds to a given atom are summed up to obtain a resulting bond vector. For example, in Figure 1, the resulting bond vector for the first nitrogen atom points directly toward the second nitrogen atom. If there are multiple bonds like in H₂O, the resulting bond vector for oxygen points in between the two H atoms. Lone FODs have to reside in

| System | $E_{\text{OPT - SCAN}}$ | $E_{\text{GUESS}}$ | $E_{\text{OPT - FLO-SIC}}$ | $\Delta E_{\text{FLO - SIC}}$ | $F_{\text{max}}$ |
|--------|-------------------------|-----------------|--------------------------|------------------|----------------|
| H      | -0.5001                 | -0.4999        | -0.4999                  | 0.0              | 0.0            |
| He     | -2.9049                 | -2.8995        | -2.8995                  | 0.0              | 0.0            |
| Li     | -7.4799                 | -7.4733        | -7.4741                  | 0.8              | 2.8            |
| Be     | -14.6496                | -14.6400       | -14.6425                 | 2.5              | 4.8            |
| B      | -24.6406                | -24.6248       | -24.6268                 | 2.0              | 0.6            |
| C      | -37.8390                | -37.8081       | -37.8104                 | 2.3              | 0.8            |
| N      | -54.5886                | -54.5299       | -54.5310                 | 1.1              | 1.7            |
| F      | -99.7447                | -99.6296       | -99.6326                 | 3.0              | 0.6            |
| BH₃    | -26.5920                | -26.5765       | -26.5766                 | 0.1              | 2.3            |
| C₂H₂   | -77.3172                | -77.2220       | -77.2225                 | 0.5              | 2.4            |
| C₂H₄   | -78.5746                | -78.4890       | -78.4896                 | 0.6              | 3.6            |
| C₂H₆   | -79.8152                | -79.7364       | -79.7366                 | 0.2              | 2.1            |
| C₃H₃   | -117.2480               | -117.1142      | -117.1177                | 3.5              | 2.9            |
| C₂H₅(Lewis) | -232.2259               | -231.9401      | -231.9425                | 2.4              | 3.7            |
| C₂H₅(Linnett) | -232.2259               | -231.9424      | -231.9448                | 2.4              | 4.3            |
| CH₃    | -39.8369                | -39.7905       | -39.7942                 | 3.7              | 4.9            |
| CH₄    | -40.5082                | -40.4681       | -40.4681                 | 0.0              | 0.5            |
| CN     | -92.7082                | -92.5654       | -92.5723                 | 6.9              | 3.5            |
| F₂     | -199.5495               | -199.2660      | -199.2833                | 17.3             | 2.0            |
| H₂O₂   | -151.5659               | -151.3495      | -151.3585                | 9.0              | 2.7            |
| HCN    | -93.4158                | -93.2891       | -93.2913                 | 2.2              | 1.8            |
| HF     | -100.4641               | -100.3233      | -100.3246                | 1.3              | 1.9            |
| N₂     | -109.5240               | -109.3675      | -109.3753                | 7.8              | 3.8            |
| N₂H    | -164.7884               | -164.5142      | -164.5209                | 15.7             | 4.8            |
| CH₃OH  | -115.7261               | -115.5796      | -115.5824                | 2.8              | 4.6            |
| C₂H₅OH | -155.0413               | -154.8553      | -154.8584                | 3.1              | 1.8            |
| CH₂F   | -139.7579               | -139.5794      | -139.5834                | 4.0              | 1.5            |
| CH₂F₂  | -239.0263               | -238.7057      | -238.7142                | 8.5              | 2.9            |
| CH₂F₃  | -338.3064               | -337.8443      | -337.8577                | 13.4             | 4.1            |
| CF₄    | -437.5731               | -436.9624      | -436.9811                | 18.7             | 3.8            |
| CH₂F   | -139.0847               | -138.9013      | -138.9087                | 7.4              | 4.4            |
| CH₂NH  | -94.5985                | -94.4856       | -94.4875                 | 1.9              | 4.2            |
| C₂H₂O  | -153.8185               | -153.6246      | -153.6296                | 5.0              | 3.7            |
| CH₂O   | -114.4851               | -114.3350      | -114.3392                | 4.2              | 3.7            |
| CH₂NO  | -169.8791               | -169.6405      | -169.6515                | 11.0             | 4.2            |
| CH₂SH  | -438.7455               | -438.3446      | -438.3497                | 5.1              | 3.5            |
| C₂H₅SH | -478.0563               | -477.6162      | -477.6208                | 4.6              | 2.2            |
the opposite direction of this vector (see Supporting Information as well). For multiple lone FODs, the same procedure as for multiple bond FODs is used. This leads to a simple and intuitive formation of bonds and lone pairs. For planar or linear molecules (other than diatomics) where atoms have lone FODs, such FODs are oriented perpendicular to the molecular plane or the molecular axis, respectively. Overall, this approach yields FOD geometries in correspondence to LDQ theory.\cite{37,38}

After all valence FODs are initialized, core FODs are introduced by symmetrically placing them on their respective shells. The radii and number of FODs for the subshells for all atoms are stored in a database file. Afterward, they are optimized with respect to the valence FODs, that is, $r_i$ is minimized globally via rigid rotations (see Fig. 3).

The fodMC is a computationally efficient method. It needs nuclear positions as well as a FOD bond order as input, which allows calculating initial FODs for any bond order. This is especially important for systems where the bonding situation is ambiguous. For instance, the fodMC allows an explicit generation of Lewis and Linnett FOD guesses, which can then be compared using FLO-SIC. Because not all bond orders that can be initialized are physically or chemically meaningful, this method should be used with care. It is not a black-box method.

As pointed out by Lehtola et al.,\cite{93} variational PZ-SIC can suffer from a multiple minima problem. This problem might be addressed, at a higher computational cost, by performing a stability analysis. In FLO-SIC, the electronic geometry built up by the FODs offers a computationally cheap alternative to analyze the resulting SIC solutions. With the fodMC, one can initialize any arrangement of FODs for a given system and judge which configuration corresponds to the most stable SIC solution.

### Python-based Lewis

Prior application of the FLO-SIC method to atoms and molecules\cite{42,45,34,35} showed that there are repeating structural motifs for the FODs describing orbitals for core and valence electrons. As discussed previously, in the case of valence electrons, these structural motifs carry bonding information and are transferable between similar chemical bonding situations.

PyLEWIS uses these known structural motifs to generate the FODs. The structural motifs are precalculated using fodMC or PyEFF and are stored as fixed data sets. The placement of the descriptors is then based on the data sets and fixed rules or geometrical constraints. This has to be obeyed. There is no further optimization of the FOD positions after placing the FODs according to the fixed data sets and rules. Due to the geometrical contraints, the generated guesses may not be as good as the guesses generated with the other methods.

For example, FODs corresponding to 1$s$ core states are located at the respective nuclear position and FODs for 2s2p orbitals are arranged as tetrahedra around the nucleus. FODs describing a single bond are located on the bond axis between the respective nuclei (one up and one down FOD), FODs corresponding to a double bond hover above and below the bond axis (one up and one down FOD for each position), and FODs representing a triple bond are placed on a triangle around the bond axis (one up and one down FOD at each vertex of the triangle). In addition, FODs representing single unpaired electrons are placed at an elongation of the bond axis at the side of the respective nuclei (one up and one down FOD), FODs representing a single unpaired electron are located on the bond axis between the respective nuclear positions as well as a FOD bond order as input, which allows calculating initial FODs for any bond order. This is especially important for systems where the bonding situation is ambiguous. For instance, the fodMC allows an explicit generation of Lewis and Linnett FOD guesses, which can then be compared using FLO-SIC. Because not all bond orders that can be initialized are physically or chemically meaningful, this method should be used with care. It is not a black-box method.

As pointed out by Lehtola et al.,\cite{93} variational PZ-SIC can suffer from a multiple minima problem. This problem might be addressed, at a higher computational cost, by performing a stability analysis. In FLO-SIC, the electronic geometry built up by the FODs offers a computationally cheap alternative to analyze the resulting SIC solutions. With the fodMC, one can initialize any arrangement of FODs for a given system and judge which configuration corresponds to the most stable SIC solution.

### Table 5. Errors in the RPDF for the initial FOD positions of the different FOD generators in comparison with the optimized ones. The comparison includes the mean error (ME, first value), mean absolute error (MAE, second value), and the root mean square deviation (RMSD, third value). The smallest errors of each type are highlighted in bold for each system.

| System   | PyEFF  | PyCOM | fodMC | PyLEWIS |
|----------|--------|-------|-------|---------|
| Be       | 0.083  | 0.111 |       | –       |
|          | 0.432  | 0.461 |       | –       |
|          | 6.962  | 7.496 |       | –       |
| B        | 0.169  | 0.041 |       | –       |
|          | 0.711  | 0.268 |       | –       |
|          | 10.949 | 4.282 |       | –       |
| C        | 0.362  | 0.013 |       | –       |
|          | 0.955  | 0.606 |       | –       |
|          | 14.047 | 7.780 |       | –       |
| N        | 0.958  | 0.035 |       | –       |
|          | 1.949  | 0.626 |       | –       |
|          | 29.395 | 9.910 |       | –       |
| F        | 3.701  | –0.016|       | –       |
|          | 5.254  | 1.397 |       | –       |
|          | 15.195 | 15.195|       | –       |
| BH₃      | 0.031  | 0.000 | 0.680 | –       |
|          | 0.731  | 1.208 |       | –       |
|          | 8.600  | 16.105|       | –       |
| C₂H₂     | -0.001 | 0.009 | 2.080 | –       |
|          | 0.924  | 2.981 |       | –       |
|          | 7.295  | 40.323|       | –       |
| C₂H₄     | 0.002  | -0.038| 0.057 |       |
|          | 0.808  | 0.732 | 0.862 |       |
|          | 5.838  | 5.626 | 5.488 |       |
| C₂H₆     | 0.010  | -0.012| 0.350 |       |
|          | 0.720  | 0.564 | 1.060 |       |
|          | 5.279  | 4.083 | 6.688 |       |
| CH₃      | 0.067  | 0.006 | 0.195 |       |
|          | 0.892  | 0.831 | 1.020 |       |
|          | 10.534 | 8.985 | 9.291 |       |
| CH₄      | 0.056  | 0.000 | 0.746 |       |
|          | 0.895  | 0.537 | 1.585 |       |
|          | 9.745  | 6.700 | 16.066|       |
| C₂H₅     | 414.499| -0.019| 0.040 |       |
|          | 415.032| 0.555 | 0.630 |       |
|          | 9268.837| 1.973 | 2.685 |       |
| CN       | -0.155 | 87.793| 0.194 | 5.087  |
|          | 1.460  | 89.407| 5.628 |       |
|          | 9.788  | 1957.828| 11.605| 117.936|
|          | 6.472  | 7787.944| 6.885 | 134.655|
| N₂       | 0.942  | 0.934 | 8.861 |       |
|          | 1.398  | 1.705 | 10.179|       |
|          | 15441.296| 5.277 | 48.153|       |
| HF       | -3.255 | 0.018 | 11.908|       |
|          | -4.885 | 1.212 | 13.539|       |
|          | -59.693| 11.522| 177.006|       |
|          | -1.157 | 0.154 | 7.013 |       |
|          | -2.889 | 1.663 | 8.522 |       |
|          | 33.466 | 13.762| 109.232|       |
|          | -0.897 | -0.061| 5.230 |       |
|          | 1.860  | 0.904 | 6.222 |       |
|          | 12.913 | 4.421 | 78.012|       |
one nuclear position. The orientation of the different structural motifs with respect to each other is based on fixed rules gained from analyzing the results of the fodMC method.

The input provided by the user is the bonding information of the molecule. The idea is to read a Lewis structural formula of the molecule and to produce a set of FODs corresponding to that input. Standard formats like the so-called SMILES string as well as the common protein database (pdb) file can be used.

The bonding information is translated into a xyz file containing nuclei and descriptor positions. PyLEWIS is written in python and uses the rdkit[96] python package.

**Benchmark Set**

To compare the performance of all four methods introduced in the previous section, we choose a small benchmark set consisting of atoms and molecules with single bonds (BH₃, CH₄, H₂O₂, HF, F₂, C₂H₆), double bonds (C₃H₅, C₂H₄), triple bonds (HCN, N₂, C₂H₂, N₃H), aromatic rings (C₆H₆), spin-polarization (CH₃, CN), and three-dimensional character (C₆₀). C₆₀ is only discussed qualitatively to reduce the computational effort.

To extend our benchmark set to a larger variety of bonded elements, we have analyzed the following molecules: CH₂F, methanol (CH₃OH), ethanol (C₂H₅OH), formaldehyde (CH₂O), acetaldehyde (C₂H₆O), formamide (CH₃NO), monofluoromethane (CH₃F), difluoromethane (CH₂F₂), trifluoromethane (CHF₃), tetrafluoromethane (CF₄), methanimine (CH₂NH), methanethiol (CH₃SH), and ethanethiol (C₂H₅SH). For this extended study, we only used PyCOM and fodMC, as they are applicable to any system. On the contrary, initial FOD positions for these structures can be difficult or impossible to obtain with the current versions of PyEFF or PyLEWIS.

**Computational details**

All FLO-SIC calculations and FOD optimizations were performed with our recently developed PyFLOSIC code. This code offers a user-friendly python implementation of FLO-SIC within the well-maintained PySCF[90] electronic structure code.

We use a fully self-consistent approach with unified Hamiltonians.[39,97] Using the HOOOV unified Hamiltonian, we get identical total energies, densities and polarizabilities in comparison to the Pederson Hamiltonian in FG-NRLMOL.[98] All results presented in this work were obtained by employing the HOOOV. It is worth mentioning that for this study, both unified Hamiltonians (H₀₀₀ and H₀₀₀) deliver nearly identical results (see Supporting Information).

All calculations were done with a density functional optimized (DFO)[85] basis set, an energy threshold of Eₘₚₚₜ = 10⁻⁶ Eₜ, and a PySCF grid level of 7. The forces on Fermi-orbital descriptors were optimized using the atomic simulation environment (ase)[99] geometry optimizer FIRE[100] until the maximum force component drops below a threshold of F_OPT < 5⋅10⁻⁴ Eₜ/a₀. Experimental geometries were taken from the CCCBDB database.[68]

Each FOD optimization was started from fodMC guesses and performed for different rungs of density functionals, namely LDA,[52] PBE,[53] and SCAN.[54] This is done to illustrate that the same set of initial FODs is optimized differently using different exchange-correlation functionals. The results are summarized in Table 2 for

---

**Figure 4.** RPDF for BH₃. The respective nuclear RPDF is given in green, the RPDF for all up-FODs is given in blue and the RPDF for all down-FODs is given in orange. The RPDFs for PyCOM and PyEFF show the same trends, as do the RPDFs for fodMC and PyLEWIS. The best agreement with the RPDF of the optimized FODs is given by fodMC and PyEFF. [Color figure can be viewed at wileyonlinelibrary.com]
Table 6. Errors in the RPDF for the initial FOD positions of PyCOM and fodMC in comparison with the optimized ones for the extended benchmark set. The comparison includes the mean error (ME, first value), mean absolute error (MAE, second value), and the root mean square deviation (RMSD, third value).

| System                  | PyCOM  | fodMC  |
|-------------------------|--------|--------|
| CH₂F                    | 1.085  | 0.125  |
|                         | 2.067  | 1.236  |
|                         | 17.225 | 7.884  |
| CH₃OH (Methanol)        | 0.558  | -0.050 |
|                         | 1.273  | 0.884  |
|                         | 9.551  | 3.911  |
| C₂H₅OH (Ethanol)        | 0.288  | -0.031 |
|                         | 0.901  | 0.624  |
|                         | 4.927  | 2.147  |
| CH₂O (Formaldehyde)     | 0.697  | 0.053  |
|                         | 1.870  | 1.127  |
|                         | 13.179 | 6.792  |
| C₆H₁₂O (Acetaldehyde)   | 0.344  | 0.012  |
|                         | 1.127  | 0.786  |
|                         | 6.071  | 3.105  |
| CH₄N (Formamide)        | 0.460  | 0.013  |
|                         | 1.299  | 0.740  |
|                         | 5.432  | 2.753  |
| CH₂F (Monofluoromethane)| 0.965  | 0.075  |
|                         | 2.123  | 1.170  |
|                         | 16.415 | 7.955  |
| CH₂F₂ (Difluoromethane) | 0.888  | 0.086  |
|                         | 1.820  | 1.106  |
|                         | 13.703 | 7.056  |
| CHF₃ (Trifluoromethane) | 0.758  | 0.071  |
|                         | 1.652  | 0.998  |
|                         | 13.099 | 6.402  |
| CF₄ (Tetrafluoromethane)| 1.183  | 0.086  |
|                         | 1.842  | 0.834  |
|                         | 17.296 | 5.233  |
| CH₂NH (Methanimine)     | 0.398  | -0.038 |
|                         | 1.343  | 0.788  |
|                         | 7.142  | 4.710  |
| CH₂SH (Methanethiol)    | 3.169  | -0.076 |
|                         | 5.571  | 1.868  |
|                         | 54.646 | 15.952 |
| C₂H₅SH (Ethanethiol)    | 31.333 | -0.054 |
|                         | 32.788 | 1.213  |
|                         | 672.977| 8.911  |
| C₆H₆ Lewis              | 0.075  | -0.012 |
|                         | 0.563  | 0.428  |
|                         | 2.811  | 2.055  |
| C₆H₆ Linnett            | 0.081  | -0.011 |
|                         | 0.550  | 0.436  |
|                         | 2.784  | 2.169  |

FLO-SIC LDA, in Table 3 for FLO-SIC PBE, and in Table 4 for FLO-SIC SCAN. All fully optimized FOD geometries recover the structural motifs of LDQ theory. The energy change during the FLO-SIC optimization is less than 25 mEh, thus the fodMC delivers excellent starting values for a FLO-SIC calculation.

It can be difficult to analyze the characteristics of the FOD positions for the different FOD generators. The radial pair distribution function (RPDF) g(r) can deliver information about the coordination sphere of specific FODs. The RPDF provides a tool to compare relative FOD positions, making it ideal for the present study. Pair-distribution functions for nuclear positions are often used to analyze molecular dynamics trajectories. Technical details about the RPDF in general are given in the Supporting Information.

Results

All FOD generator methods were used to obtain starting FOD positions for the systems in the small test set, while for the extended test set only fodMC and PyCOM were used. These starting positions are compared to fully optimized FOD positions in FLO-SIC LDA (starting from fodMC guesses, see Table 5). Typical errors like mean error, mean absolute error, and the root mean square deviation are used to quantify the RPDF results. A small error between the RPDFs of a guess compared to the optimized FODs indicates that the guess is already close to the minimum configuration, while large errors indicate a larger deviation between the structures. Errors in the RPDF might be slightly biased due to the reference used, but this should not strongly affect the results.

As an example, we present the RPDF for the BH₃ molecule (see Fig. 4). The RPDFs for PyCOM and PyEFF describe the same trends as do the RPDFs for fodMC and PyLEWIS. The best agreement with the RPDF of the optimized FODs is given by fodMC and PyEFF.

As can be seen from Table 5, for small atoms (Be—C) and simple hydrocarbon molecules (CH₃, CH₄, C₂H₅, C₃H₆, and C₂H₄), the fodMC and PyEFF methods generate better guesses than PyCOM. The intrinsic limitation of PyEFF can easily be observed in the large errors for N₃H and C₃H₅. The errors for PyLEWIS guesses are not large, but the method needs further improvements before it can achieve an accuracy similar to that of PyCOM or fodMC. Considering the small benchmark set, the fodMC outperforms all other methods. Unlike PyEFF and PyLEWIS, fodMC and PyCOM are able to deliver guesses for any systems. Thus, we recommend using fodMC and PyCOM for systems with unknown or more complex bonding situations, such as the extended benchmark set in Table 6. The additional RPDF analysis for this benchmark set confirms that fodMC performs...
better than PyCOM, but both methods deliver good starting points even for complex structures.

**Efficiency analysis**

All FOD generator methods were compared as stand-alone tools. In the following, these tools are compared with respect to specific properties (Automation, Speed, Extensibility, Insensitivity, and Scope).

**Automation.** The input level for the PyCOM method is the lowest of all methods. The performance of PyEFF critically depends on the effective electron radii. In case of fodMC and PyLEWIS, the complete bonding matrix is needed as input.

**Scope.** Both PyCOM as well as fodMC can treat any system, whereas PyEFF (until now) only works properly for systems containing atoms from hydrogen to nitrogen. PyLEWIS only works for systems up to Ar. Thus, both PyCOM and fodMC have an unlimited application scope.

**Speed.** PyCOM needs a standard density functional theory calculation as input and it requires the numerical effort of a localization procedure, making it the computationally most expensive method. The second highest numerical effort is required by the PyEFF method, because force field calculations need to be performed. The numerically cheapest methods are fodMC and PyLEWIS. Both methods have practically no numerical cost.

**Extensibility.** Besides the starting guess, PyEFF delivers an energetical description and accordingly has utility beyond FOD generation. PyLEWIS can start from various starting information (e.g., SMILES string or pdb files), and the method can be optimized using additional constraints. Knowing that FODs are not explicitly the centroids of localized orbitals, one can optimize PyCOM by using, for example, a least square fit between pre-localized orbitals (like Foster-Boys) and the target Fermi orbitals by varying the FOD positions until the difference is minimized. This should lead to improved initial FODs. The fodMC can only be improved with respect to its FOD generation. For example, there might be better starting values to initialize the bond/lone/core FODs. Furthermore, one could introduce a Metropolis algorithm for the movement of the FODs, allowing $f_i$ to go uphill.

**Insensitivity.** The fodMC delivers one set of FOD positions for a molecular geometry and a given bonding matrix. There is no further information needed from the user. If the user changes the bonding matrix, the resulting set of FODs will change too. Similarly, changing the effective electron radii in PyEFF will result in different FODs. In the case of PyLEWIS, the user has to provide information on the molecular structure, chemical bonding, and eventually additional constraints to be applied in the FOD placement. Providing different bonding information will give different sets of FODs. The method with the worst score in...
terms of insensitivity is PyCOM, because it requires the molecular structure, all details of the DFT method and the localization method used in generating single particle densities. The resulting set of FODs will depend on the exchange-correlation functional as well as on the numerical details.

A visual efficiency analysis is given in Figure 5. This figure shows the relative performance of each method considering several aspects. All facts presented in this paragraph are related to essential properties (Automation, Scope, Speed, Extensibility, and Insensitivity) on a relative scale to analyze the strengths and weaknesses of each presented method.

Conclusions

The most significant conclusion from the work presented here is that the FODs in the FLO-SIC method carry bonding information based on their relationship to the exchange hole (and possibly the exchange-correlation hole). A FLO-SIC bond order based on FOD positions is introduced and benchmarked for structures of the G2-1 benchmark set. Notably, it was shown that fully optimized FOD geometries correspond remarkably well to LDQ theory, which is especially relevant for molecules where Lewis theory does not provide a well-defined description of the bonding.

Four different methods (PyEFF, PyCOM, fodMC, and PyLEWIS) to generate initial FOD positions for a FLO-SIC calculation are presented. These methods are compared with respect to the use of bonding information for the generation of initial FODs. We introduce new benchmark sets of molecules to test the utility of the initial FOD positions and give a detailed comparison of the FOD positions, detailing a comparison of the efficiency, the usage, the advantages, and disadvantages of the methods. Finally, we recommend checking any set of optimized FODs of a determined FLO-SIC minimum against LDQ or Lewis structures. Doing so allows the verification of the respective FLO-SIC solution because, as we have suggested here, the FODs should agree with at least one of these structures.

All codes discussed here are available under Open Source licenses (APACHE 2.0) at https://github.com/pyflosic. Further, all optimized FOD configurations are available at https://github.com/pyflosic/FODs_database.

Outlook

As discussed above, optimized FOD positions resemble LDQ structures. This is not only true for atoms and small molecules, but also for aromatic systems like the well-known benzene molecule (see Fig. 6). The corresponding FLO-SIC calculations show that the LDQ structure is energetically preferred compared to the Lewis structure. This is true for the total as well as the SIC energy (see Supporting Information).

The methods explained in this article allow the investigation of large molecules, because initial FOD positions can be obtained easily. We present initial FOD positions for molecules of chemical and biological interest, such as C60 (see Fig. 7) or vitamins like vitamin C and E (see Fig. 8). In addition, the analysis of the FOD positions opens the way to analyze chemical reactions electron-by-electron. This opens the possibility to study coupling and reaction paths in the future.

Further, we plan to re-parametrize the eFF with FLO-SIC reference data, given the promising results of PyEFF for small atoms and molecules. This may open a way for establishing embedding techniques, where only the active center is calculated with FLO-SIC and the inactive parts are calculated with a FLO-SIC optimized electronic force field. Such techniques would enable the treatment of larger macromolecules or biological systems (e.g., endohedral fullerenes, vitamins, or even DNA).

We discussed above the chemical relevance of the FODs to describe and analyze bonding and proposed several methods to find them. The proposed methods are a first step in understanding FLO-SIC and FODs better. In the future, other methods like machine learning procedures could replace these methods. Finally, all methods might be combined in one user-friendly umbrella FOD generator code as suggested by one referee of this article.

While this article addresses the chemical interpretation and generation of FODs, it also raises several new questions for the community. For example, can the FOD positions be an intrinsic descriptor for different levels of theory (wave function theories, density functional theory, and classical methods)? Can an efficient multilevel theory be formulated based on them? We hope that the contents of this article help to motivate the community to work on these important unanswered questions in order to efficiently treat physical, chemical, and biological problems computationally in the future.

Acknowledgment

S. Schwalbe and K. Trepte designed the work and wrote the first draft. S. Schwalbe coded all python based generators, while the fodMC was coded by K. Trepte. The other authors contributed ideas for the code development, code testing, discussions, and revisions. The authors thank the ZIH in Dresden for computational time. K. Trepte, A. I. Johnson, J. E. Peralta and K. A. Jackson are supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Computational Chemical Sciences Program under Award Number #DE-SC0018331. J. Kortus thanks the German Research Foundation (DFG) for financial support of this study KO 1924/9-1. The authors are grateful to our referees for their useful comments, which have increased the scientific content and quality of the manuscript. Furthermore, we would like to thank Mark R. Pederson for creating the FLO-SIC method, an achievement that has initialized our interest in the topic and started this work.

Keywords: density functional theory · FLO-SIC · chemical bonding · Linnett double-quartet theory

How to cite this article: S. Schwalbe, K. Trepte, L. Fiedler, A. I. Johnson, J. Kraus, T. Hahn, J. E. Peralta, K. A. Jackson, J. Kortus. J. Comput. Chem. 2019, 40, 2843–2857. DOI: 10.1002/jcc.26062

Additional Supporting Information may be found in the online version of this article.
[93] S. Lehtola, M. Head-Gordon, H. Jónsson, *J. Chem. Theory Comput.* 2016, 12, 3195.
[94] T. Hahn, S. Liebing, J. Kortus, M. R. Pederson, *J. Chem. Phys.* 2015, 143, 224104.
[95] D.-y. Kao, K. Withanage, T. Hahn, J. Batool, J. Kortus, K. Jackson, *J. Chem. Phys.* 2017, 147, 164107.
[96] G. Landrum, *Rdkit: Open-Source Cheminformatics.* http://www.rdkit.org.
[97] S. Lehtola, H. Jónsson, *J. Chem. Theory Comput.* 2014, 10, 5324.
[98] S. Schwalbe, L. Fiedler, K. Trepte, T. Hahn, J. Kortus, and J. Perdew, Self-Interaction Corrected Density Functional Theory is Not Straying from the Path Towards the Exact Functional, Manuscript in Preparation. 2019.
[99] Atomic Simulation Environment, 2018. https://wiki.fysik.dtu.dk/ase/.
[100] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, P. Gumbsch, *Phys. Rev. Lett.* 2006, 97, 170201.
[101] W. F. Luder, *J. Chem. Educ.* 1966, 43, 55.
[102] W. F. Luder, *J. Chem. Educ.* 1967, 44, 206.
[103] S. Schwalbe, T. Gruber, K. Trepte, F. Taubert, F. Mertens, J. Kortus, *Comput. Mater. Sci.* 2017, 134, 48.
[104] F. Taubert, S. Schwalbe, J. Seidel, R. Huettel, T. Gruber, R. Janot, M. Bobnar, R. Gumieniuk, F. Mertens, J. Kortus, *Int. J. Mater. Res.* 2017, 108, 942.

Received: 19 February 2019
Revised: 13 August 2019
Accepted: 14 August 2019
Published online on 10 September 2019