Electrons usually are paired—thereby forming a closed-shell singlet state—LT is not able to describe molecules with non-singlet ground states, such as the oxygen molecule. This shortcoming of LT, the lack of electronic spin, is solved by Linnett’s double-quartet (LDQ) theory.[3] The PZ functional explicitly removes the self-interaction error from the Kohn–Sham (KS) energy functional.[4]

\[
E^{\text{KS}}[\rho^\alpha, \rho^\beta] = T_\rho[\rho^\alpha, \rho^\beta] + V[\rho] + E_J[\rho] + E_{\text{xc}}[\rho^\alpha, \rho^\beta],
\]

where \( T_\rho[\rho^\alpha, \rho^\beta] \) is the kinetic energy of the non-interacting system, \( V[\rho] \) is the external potential energy, and \( \rho^\sigma \) is the electron density for spin \( \sigma \), with \( \alpha \) and \( \beta \) representing spin up and spin down, respectively. The PZ functional explicitly removes the self-interaction error one electron at a time.[4]

\[
E^{\text{PZ}} = E^{\text{KS}} - \sum_{i\sigma} \left( E_J[n_{i\sigma}] + E_{\text{xc}}[n_{i\sigma}] \right),
\]

where \( n_{i\sigma} \) is the electron density of the \( i \)-th occupied orbital with spin \( \sigma \), and \( E_J \) and \( E_{\text{xc}} \) denote the Coulomb and exchange-correlation energy functionals.

While the usual formulation of DFT is invariant to rotations between occupied orbitals,[5] this invariance is broken by SIC which explicitly depends on the occupied orbitals used to evaluate Eq. (2).[6] The PZ functional then needs to be minimized in terms of the \( O(N_{\text{occ}}^2) \) rotation angles for \( N_{\text{occ}} \) occupied spatial orbitals; the orbitals tend to become localized upon the optimization.[6] It has been shown more recently that the variational optimization requires complex orbitals,[7] software for such calculations is openly available.[8]
In the Fermi–Löwdin orbital (FLO) formulation of SIC, the need for occupied-occupied orbital rotations is avoided by the use of Fermi orbitals, obtained from for spin-$\sigma$ position eigenstates $|a^\sigma_i\rangle$ localized at the so-called Fermi-orbital descriptor (FOD) $a^\sigma_i$ which become the optimized variables in the theory. When the occupied orbitals are known as a basis set expansion

$$\psi^\sigma_j(r) = \sum_\mu C^\sigma_{ji} \chi^\mu_j(r), \quad (3)$$

where $C^\sigma$ are the molecular orbital coefficients and $\chi^\mu_j$ are the basis functions, the corresponding Fermi orbital (FO) coefficients $c^\sigma_{FO}$ are given by

$$c^\sigma_{FO} = C^\sigma R^\sigma, \quad (4)$$

where the transformation matrix $R^\sigma$ is given by

$$R^\sigma_{ji} = \frac{\langle \psi^\sigma_j | a^\sigma_i \rangle}{\sqrt{n^\sigma(a^\sigma_i)}}. \quad (5)$$

The FOs are non-orthonormal but can be symmetrically orthogonalized to yield the Fermi–Löwdin orbitals (FLOs) with coefficients

$$c^\sigma = c^\sigma_{FO} [T^\sigma (Q^\sigma)^{-1/2} (T^\sigma)^T], \quad (6)$$

where $T^\sigma$ and $Q^\sigma$ contain the eigenvectors and eigenvalues of the FO overlap matrix, respectively. These FLOs are then used to evaluate the PZ functional, Eq. [2]. In the spin-restricted formalism, the Fermi orbitals are parametrized by a set of $3N_{occ}$ Fermi orbital descriptors (FODs), which correspond to semiclassical cartesian coordinates of the $N_{occ}$ electrons of each spin that need to be optimized instead of the $O(N^2_{occ})$ rotation angles of the traditional formalism. Analytical energy gradients for the FODs have been derived in Ref. [13] and they are employed in various implementations of the FLO-SIC method; please refer to Schwalbe et al. for a thorough overview on the presently used, openly available implementation of FLO-SIC.

Schwalbe et al. showed that the FODs and the corresponding FLOs have a clear relation to chemical bonding theories. The most common bond patterns in LT are single, double, and triple bonds, and any of these bonds can be represented in terms of FOD arrangements. For instance, single bonds are described by one $\alpha$ and one $\beta$ FOD in-between the atoms, double bonds are characterized by two $\alpha$ and two $\beta$ FODs above and below the bond center, and triple bonds are built up by three $\alpha$ and three $\beta$ FODs placed in triangles around the bond center for each spin. Analogous FOD motifs have also been found for FODs corresponding to core and lone pair orbitals, for example, FODs corresponding to 1s orbitals are typically placed at the nuclear sites, while the FODs for fully occupied 2s and 2p orbitals usually form a tetrahedron.

While LDQ theory includes all the bonding motifs of LT, it also allows a simple and elegant representation of aromatic bonds. As illustrated by Fig. 1, an aromatic bond can be represented in LDQ by two $\alpha$ FODs and one $\beta$ FOD or vice versa, leading to a formal bond order of 1.5, whereas the LT structures have alternating single and double bonds.

![Figure 1](image.png)

In this work, we investigate these findings further by studying the usefulness of FODs that follow the bonding theories of Lewis and Linnett for exploring symmetry breaking in SIC calculations. Symmetry breaking in SIC was originally discovered by Lehtola, Head-Gordon, and Jönsson, who showed that the molecular symmetry of benzene is broken by SIC with real (RSIC) or complex orbitals (CSIC). As discussed in Ref. [15] the symmetry breaking arises because SIC calculations on benzene are based on a single LT structure with alternating single and double bonds. The energy of the SIC solution can be decreased by allowing the C-C double bonds in the Lewis structure to contract, leading to more spatial localization which is energetically favored by SIC, while the C-C single bonds of the Lewis structure become elongated. The carbon-carbon bonds in benzene then come out unequal in length, similarly to (but not as strongly as in) the fictitious 1,3,5-cyclohexatriene molecule.

While SIC solutions guided by LT thus break the molecular symmetry of benzene, we show in this work that LDQ theory succeeds in reproducing a symmetric molecular geometry and thereby overcomes symmetry breaking for this molecule. We also show that the symmetry breaking of LT can already be seen in the electron density at the symmetric molecular geometry of benzene.
To highlight the broader scope of these findings for various kinds of SIC, in this work we consider not only the FLO-SIC method, but also include variational SIC in both the RSIC and CSIC variants to show how the choice of the initial orbital guess affects self-consistent SIC solutions.

Moreover, as artifacts arising from the use of a single Lewis structure may also become visible in the electron density, we point out the need to carefully analyze SIC solutions. Following the pioneering work of Hait and Head-Gordon, who suggested using electronic dipole moments (DMs) as a simple global measure of the electron density reproduced by DFT, we investigate with DMs whether the electron density from SIC calculations properly reflects the molecular symmetry. Using a benchmark set of 16 planar cyclic molecules, which extends a database previously suggested by Adhikari et al., we show that bonding patterns guided by LT lead in some cases to SIC DMs that do not reflect the molecular symmetry, leading to artificial dipole moments. For example, in the case of phenazine, \((C_6H_4)_2N_2\), which is symmetric and therefore should not have a dipole moment, LT structures result in a huge artificial dipole moment of 1 Debye.

Motivated by the success with benzene, we also show how LDQ can be used to find alternative bonding patterns that lead to improved DMs in SIC calculations for this database. The LDQ structures for all the studied molecules have the same symmetries as the molecules themselves, at variance to the LT structures. The correct symmetry of the FODs also leads to DMs with the correct symmetry; in the previously discussed case of phenazine, LDQ correctly yields no dipole moment. Based on this victory, we point out that LDQ FODs can be seen as a simple way to model aromatic molecules in SIC. We elaborate that FLOs based on LDQ FODs, which can be interpreted as a superposition of LT structures, can be used as initial guesses for RSIC and CSIC, leading to stable local minima.

This work is structured as follows. The methodology used in this study is discussed in Section II and details of the electronic structure calculations carried out in this work are provided in Section III. The results are presented in Section IV. First, various local minima and symmetry breaking in benzene are discussed in detail using FLO-SIC, RSIC, and CSIC in Section IV.A. Next, the influence of the initial orbital guesses following LT or LDQ arrangements on the DMs arising from SIC calculations on our benchmark set of planar, cyclic molecules is discussed in Section IV.B, where calculated DMs are used to explain how the SIC solutions are affected by the chemical bonding situations described by the initial orbital guess; basis set convergence is also investigated. Further investigations into bonding motifs are discussed in Section IV.C. The main findings of this study are summarized and conclusions are drawn in Section V.

II. METHODS

A. Benzene study

Our goal in this section is to study the effect of various electronic geometries, that is, FOD arrangements, on the molecular structure of benzene. Benzene is the prototypical aromatic molecule, and as will be shown later in this work, any conclusions that can be derived for benzene are likely applicable also to larger aromatic systems as well. As has been already established, SIC prefers a symmetry-broken structure for benzene; the energy landscape of benzene was therefore studied as follows. Average C-C bond lengths \(d_{CC}\) in the range between 1.317 Å and 1.417 Å were sampled with a step size of 0.005 Å, yielding 21 values for the average bond length; experimentally, all the C-C bond lengths are the same, \(d_{CC} = 1.397\) Å. All C-H bonds were fixed to their experimental value, \(d_{CH} = 1.084\) Å. To investigate the distorted structures that may be favored by SIC, following Ref. [15] we elongate three C-C bonds by an amount \(\Delta\), whereas the other three C-C bonds are shortened; the hydrogens are moved correspondingly. We consider local distortions \(\Delta\) ranging from 0 Å (symmetric structure) to 0.075 Å, likewise sampled in 0.005 Å steps, leading to 16 distorted structures for each value of the average C-C bond length and a grand total of 336 candidate structures. A detailed description of how to obtain the nuclear coordinates with this procedure is given in the supporting information (SI); for a visual explanation, see Fig. 2.

The three FOD starting structures (LT1, LT2, and LDQ) shown in Fig. 5 were considered for benzene. In the LT1 structure, the double-bond FODs are placed between \(C_2-C_3\), \(C_4-C_5\) and \(C_1-C_6\); the numbering is shown in Fig. 2. In the LT2 structure, the double bonds are placed between \(C_1-C_2\), \(C_3-C_4\) and \(C_5-C_6\), instead. The LDQ structure is staggered: a double-bond FOD structure is picked for one spin channel, while a single-bond FOD structure is used for the other spin channel, and the configurations alternate spins between each bond. The LDQ structure thereby formally corresponds to a broken-symmetry open-shell singlet state; however, it correctly yields a formal bond order of 1.5 for all C-C bonds which are expected to be aromatic.

B. DIP16 benchmark set

The benchmark set used within this study extends an existing benchmark set introduced by Adhikari et al., a full list of the molecules is given in Table I along with the corresponding shorthand identifiers used in this work. The original benchmark set of Ref. [25] consists of 14 molecules containing 44 to 120 electrons and H, C, N, O, F, S, and Cl atoms. All systems in the set are cyclic and can be divided into aromatic and non-aromatic molecules. Adhikari et al. only compared the DMs of 10 of their 14 molecules to exper-
**Table 1. Description of the DIP16 benchmark set, including the molecules’ identifiers (IDs) and point groups (PGs). More details can be found in the SI.**

|Molecule | Formula | Group |
|---------|---------|-------|
| mol01   | C<sub>2v</sub> acridine | C<sub>13</sub>H<sub>9</sub>N | aromatic |
| mol02   | D<sub>2h</sub> anthracene | C<sub>14</sub>H<sub>10</sub> | aromatic |
| mol03   | C<sub>2v</sub> azulene | C<sub>10</sub>H<sub>8</sub> | aromatic |
| mol04   | D<sub>2h</sub> p-benzoquinone | C<sub>4</sub>H<sub>4</sub>O<sub>2</sub> | non-aromatic |
| mol05<sup>a</sup> | C<sub>2v</sub> 2,1,3-benzothiadiazole | C<sub>6</sub>H<sub>5</sub>N<sub>2</sub>S | aromatic |
| mol06   | C<sub>6v</sub> benzothiazole | C<sub>6</sub>H<sub>5</sub>NS | aromatic |
| mol07   | D<sub>2h</sub> p-chloranil | C<sub>6</sub>Cl<sub>4</sub>O<sub>2</sub> | non-aromatic |
| mol08<sup>a</sup> | C<sub>2v</sub> dichlorine | C<sub>10</sub>H<sub>4</sub>Cl<sub>2</sub>O<sub>2</sub> | non-aromatic |
| mol09   | D<sub>2h</sub> p-fluoranil | C<sub>6</sub>F<sub>4</sub>O<sub>2</sub> | non-aromatic |
| mol10<sup>b</sup> | C<sub>2v</sub> fluorene | C<sub>13</sub>H<sub>10</sub> | non-aromatic |
| mol11<sup>b</sup> | C<sub>2v</sub> 1,4-napthalenedione | C<sub>10</sub>H<sub>6</sub>O<sub>2</sub> | non-aromatic |
| mol12   | D<sub>2h</sub> phenazine | C<sub>12</sub>H<sub>8</sub>N<sub>2</sub> | aromatic |
| mol13<sup>a</sup> | C<sub>2v</sub> 1,2,5-thiadiazole | C<sub>5</sub>H<sub>5</sub>N<sub>2</sub>S | aromatic |
| mol14   | C<sub>2v</sub> thiophene | C<sub>4</sub>H<sub>4</sub>S | aromatic |
| mol15<sup>b</sup> | D<sub>2h</sub> cyclobutadiene | C<sub>4</sub>H<sub>4</sub> | anti-aromatic |
| mol16<sup>b</sup> | D<sub>6h</sub> benzene | C<sub>6</sub>H<sub>6</sub> | aromatic |

For the entire DIP16 test set, we investigated a total of 51 LT and LDQ electronic geometries as initial starting points for RSIC, CSIC, and FLO-SIC calculations. There are up to 4 LT and 2 LDQ electronic geometries for each molecule; an example will be provided in Section IV, and some more visualizations are available in the SI. The various LT and LDQ configurations for each molecule lead to distinct DMs, as the total density that arises from the orbital optimization is biased by the bonding pattern described by the FOD configuration. The DM components of DFT, RSIC, CSIC, and FLO-SIC calculations are also detailed in the SI.

**III. COMPUTATIONAL DETAILS**

**A. Electronic structure codes**

The Gaussian-basis all-electron codes ERKALE<sup>15,16,30,31</sup>, PySCF<sup>32</sup>, and PyFLOSIC<sup>23</sup> were used in this work. As ERKALE and PyFLOSIC are based on similar computational approaches (both employ the standard approach used in almost all Gaussian-basis quantum chemistry programs, which allows routinely reproducing total energies to micro-hartree accuracy), they enable comparisons of FLO-SIC, RSIC, and CSIC on an equal footing. The codes used in this work have a focus on simple input and output file handling as well as user-friendliness, which makes setting up complex workflows rather easy. Moreover, all...
three codes are freely and openly available, and can be used without limitations also in industry; see Ref. 33 for thorough discussion.

Furthermore, all three codes evaluate density functionals with the open-source Libxc library,33 again ensuring comparability of the results. The Perdew–Wang (PW92) correlation functional,35 which is a local density approximation (LDA), has been used for all calculations described in this work in combination with LDA exchange.25,26 if not stated otherwise.

Even though the FODs depend on the exchange-correlation functional, in analogy to the findings of Ref. 15, the results obtained at the LDA level of theory with the PW92 functional suffice to demonstrate symmetry-breaking effects arising from the self-interaction correction. Indeed, as will be demonstrated in Section IVB5, calculations performed at the generalized gradient approximation (GGA) as well as the meta-GGA levels of theory yield results that are similar to those obtained at the LDA level with the PW92 functional.

The all-electron calculations in ERKALE, PySCF and PyFLOSIC were carried out with the double-ζ polarization consistent pc-1 basis set.25,26 While DMs are a textbook example of the need for diffuse basis functions,26 as the expectation value of the dipole moment operator \( \langle \hat{\mathbf{r}} \rangle \) may accumulate significant contributions from small changes in the electron density at large \( \mathbf{r} \) where only diffuse functions contribute, the pc-1 basis set was chosen for comparability to the work of Adhikari et al.25 that employed the DFO basis set,24 which is limited to \( d \) functions, thereby contains only a single shell of polarization functions on main group elements, and is therefore formally of polarized double-ζ quality; this is also supported by numerical evidence.25 Although the use of such small basis sets is generally not advisable for investigating properties as challenging as DMs, exploratory calculations carried out with larger basis sets in Section IVB5 show that the findings of the present study are not compromised by the use of such a small basis set that was chosen to optimize the computational cost.

1. **PySCF and PyFLOSIC**

A grid level24 of 7 was used in PySCF and PyFLOSIC. This corresponds to using a (90,974) grid for hydrogen, a (135,1202) grid for second-period atoms and a (140,1202) grid for third-period atoms in Becke’s multicenter quadrature scheme.25 In contrast to the total electron density in normal DFT calculations, the orbital densities considered in SIC are not smooth; this is the main reason for the need for huge integration grids in SIC calculations, and also the rationale for turning off grid pruning in all DFT and SIC calculations of this work. Any linear dependencies in the molecular atomic-orbital basis set were removed by applying the `scfaddons.remove_linear_dep_` function on the PySCF DFT object; default PySCF thresholds for this procedure were used. An SCF energy criterion of \( 10^{-8} \, E_h \) was used in all calculations.

Starting values for the FODs for PyFLOSIC calculations were generated with the Fermi-orbital descriptor Monte-Carlo fodMC code (https://github.com/pyflosic/fodMC). While there are other ways to initialize the FODs (see Refs. 9 and 23 for discussion), the fodMC code enables a direct translation between specific chemical bonding situations and FOD guesses, making FODMC the ideal approach to study the differences between LT and LDQ within SIC. The starting FODs were optimized in the FLO-SIC calculations performed with PyFLOSIC until the maximum force max[\( |F_i| \)] acting on any FOD fell below \( 5 \cdot 10^{-4} \, E_h/a_0 \).

As it is well-known that a vanishing gradient does not necessarily imply convergence onto a local minimum, for verification purposes, we introduce a stability analysis of FLO-SIC solutions in analogy to the method introduced in Ref. 15 for RSIC and CSIC calculations. A FLO-SIC calculation will be deemed to have converged onto a stable solution if the FOD Hessian is positive semidefinite. In this work, the FOD Hessian is formed semianalytically via finite differences of analytical FOD forces, using a two-point stencil with a step size of 0.001 Å. Because of the small number of FODs, the FOD Hessian fits in memory even for large systems and can be diagonalized using dense matrix algebra.

2. **ERKALE**

An unpruned (150,1202) quadrature grid was used in ERKALE. The threshold for the gradients of the orbital rotations in the occupied-occupied block was set to \( 10^{-5} \, E_h \).

Stability analyses—checks that the SIC solutions correspond to local minima with respect to rotations of the occupied orbitals—were carried out for all RSIC and CSIC calculations; the CSIC calculations were started from stable RSIC minima according to the best practices established in Ref. 15.

The default initial orbital guess in ERKALE for RSIC and CSIC calculations is given by localized orbital24 according to the FB criterion. To allow starting RSIC calculations from the same initial guess as PyFLOSIC, ERKALE was extended in this work to be able to initialize RSIC and CSIC calculations from FODs. Thus, instead of performing a FB localization to get initial orbitals for SIC calculations, input FODs are read in alongside a converged DFT wave function; FLOs are then constructed from these data using Eqs. (4)–(6). After the initialization, the Perdew–Zunger functional, Eq. (2), is minimized as usual by variational optimization, using the orbital rotation technique detailed in Ref. 15.
IV. RESULTS

A. Benzene study

As was discussed in Section II, benzene has three distinct FOD geometries: LT1, LT2, and LDQ, which were already shown in Fig. 4. Because LT1 and LT2 only differ by a rotation of the molecule, it is expected that both LT configurations converge onto equivalent minima. The LT structures are expected to yield symmetry-broken geometries, such that the double-bond FODs prefer shorter C-C bonds. On the other hand, as the LDQ configuration can be thought of as a superposition of the two LT configurations, in which the spin-up FODs are picked from one LT structure while the spin-down FODs are picked from the other LT structure, the LDQ structure is expected to prefer a symmetric molecular geometry.

| Code | Method | Bond length [Å] | Total energy [E_h] | S^2 |
|------|--------|----------------|-------------------|------|
| PySCF | DFT    | 1.392         | 1.392 -230.042959 | 0.000 |
| PyFLOSIC | LT1  | 1.322         | 1.405 -233.023405 | 0.000 |
|        | LT2  | 1.322         | 1.405 -233.023405 | 0.000 |
|        | LDQ  | 1.362         | 1.362 -233.017015 | 0.150 |
| PySCF | FLO-SIC | LT1    | 1.392         | 1.392 -230.042959 | 0.000 |
|        | RSIC  | FB           | 1.322         | 1.405 -233.031182 | 0.000 |
|        |       | LT1          | 1.322         | 1.405 -233.031182 | 0.000 |
|        |       | LT2          | 1.322         | 1.405 -233.031182 | 0.000 |
|        |       | LDQ          | 1.362         | 1.362 -233.024006 | 0.148 |
| ERKALE | CSIC | FB           | 1.337         | 1.390 -233.082400 | 0.000 |
|        |       | LT1          | 1.337         | 1.390 -233.082394 | 0.000 |
|        |       | LT2          | 1.337         | 1.390 -233.082394 | 0.000 |
|        |       | LDQ          | 1.362         | 1.362 -233.079975 | 0.098 |

The bond lengths and total energies for DFT, RSIC, CSIC, and FLO-SIC are shown in Table 2. While DFT predicts a symmetric molecular structure, RSIC and CSIC break the symmetry as found by Lehtola, Head-Gordon, and Jónsson[23]. For RSIC, we find C-C bond lengths of 1.322 Å and 1.405 Å, while for CSIC the bond lengths are 1.337 Å and 1.390 Å. As expected, these results are in excellent agreement with the ones of Ref. 15 that found bond lengths of 1.32 Å and 1.40 Å for RSIC and 1.33 Å and 1.39 Å for CSIC with the cc-pVTZ basis set, respectively. This result serves as a consistency check for our further discussion. Furthermore, the DFT energies produced by PySCF and ERKALE are identical, showing that the calculations performed with the two codes can be meaningfully compared.

Moving onto FLO-SIC, the calculations with PyFLOSIC started from LT type guesses give analogous results to the RSIC calculations with ERKALE: the molecular structure is distorted, as the minimal-energy structure corresponds to short double bonds and long single bonds. Although FLO-SIC expectedly produces a higher total energy than RSIC, as the latter method has more optimizable degrees of freedom as it allows all possible real-valued occupied-occupied orbital rotations, the optimal bond lengths from RSIC and FLO-SIC match to the precision used in this study.

In contrast, the electronic geometry from DFT leads to a symmetric molecular structure. The solution obtained from the LDQ initial guess is slightly higher in energy (by about 6 mE_h) than that from the LT initial guesses in the RSIC calculations in ERKALE; a similar energy difference is also observed in the FLO-SIC calculations in PyFLOSIC. The energy difference of the LT and LDQ solutions is smaller in the CSIC calculations performed with ERKALE, measuring about 2.5 mE_h. However, these millihartree differences are negligible compared to the overall magnitude of the SIC, which is in the order of 3 E_h as evidenced by the data in Table 2.

The C-C bond length optimized with DFT and the PW92 functional is in excellent agreement with the experimental bond length 1.397 Å, with a difference of only 0.005 Å, which is also the resolution of this study. The LDQ calculation underestimates the experimental bond length by about 0.035 Å, whereas the LT structures break the symmetry, predicting one bond to be 0.075 Å shorter and the other to be 0.008 Å longer than in the experiment. As these results show, not only does LDQ resolve the symmetry breaking, but the maximal error in the bond lengths is twice smaller than in the LT structures. Interestingly, the optimal bond length for the LDQ structure, 1.362 Å, is close to the average of the optimal bond lengths for the LT structures, 1.364 Å. As can be expected from the correspondence of the LDQ structure to broken-symmetry open-shell singlet calculations, the spin symmetry is slightly broken as evidenced by the non-zero (S^2) from the LDQ calculations using RSIC, CSIC and FLO-SIC. In contrast, there is no spin contamination in the LT structures, but they do break the molecular symmetry.

The electron densities resulting from DFT and the RSIC solutions guided by the LT1, LT2 and LDQ FOD starting guesses in ERKALE are shown in Fig. 4. As expected, DFT delivers a symmetric electron density. Strikingly, symmetry breaking is already observable in the electron density of the two LT solutions at the symmetric molecular structure reproduced by DFT (see Table 2); both LT1 and LT2 break the rotational symmetry by placing more electron density onto the double bonds of the corresponding Lewis structure.

LDQ restores a rotationally invariant electron density, like the one from DFT. However, as expected from the open-shell singlet character of LDQ evidenced by the the non-zero value of (S^2), the LDQ calculation results
Figure 4. Total electron density in benzene according to DFT as well as RSIC initialized with FLOs following an LT or LDQ structure. Subfigures (a–d) show a wireframe representation of the densities with an isosurface value of 0.004 electrons/Å^3. Subfigures (e–h) show the corresponding two-dimensional contour plots of the density at z = 0.4 Å above the molecular plane.

Figure 5. Spin density of benzene from the LDQ solution. a) Isosurfaces of the spin density from two perspectives. b) Two-dimensional contour plots of the density at z = 0.4 Å above the molecular plane.

in a spin density, which is shown in Fig. 5 the spin-alternation of the single and double bond FODs in the LDQ structure (see Fig. 3) shows up as an alternating spin density.

B. DIP16

Having exemplified the existence of several possible types of SIC solutions for benzene, we will next examine the DMs arising from various types of SIC solutions for our benchmark database of 16 planar cyclic molecules. In the interest of keeping the discussion short, the large set of data is only discussed summarily in the main text; thorough analyses of all molecules can be found in the SI. Note again that as discussed in Section IV A all calculations in this section are performed at fixed geometries, at variance to the calculations on benzene discussed in Section IVA.

1. Aromatic molecules

Fig. 6 shows the DMs of azulene (mol03: C_{10}H_{8}, C_{2v} point group) corresponding to DFT and stable RSIC solutions from ERKALE. The calculations for Fig. 6 were started from FLOs generated by three distinct sets of FODs, analogously to the ones for benzene discussed in Section IVA. The three FOD starting points (LT1, LT2, and LDQ) each lead to a stable solution in ERKALE, as verified by the stability analysis with respect to rotations of the occupied orbitals discussed in Section III A 2. Stability analysis was also employed in the FLO-SIC calculations on this molecule (see Section III A 1), and the three solutions were each found to be true local minima.

As can be seen from Fig. 6 the DM from SIC started from an LDQ configuration points along the two-fold symmetry axis; this means that the LDQ structure leads to an electron density that reflects the symmetry of the molecule. In contrast, the electron densities obtained with SIC starting from LT structures violate this symmetry, and as a result the DMs do not point in the direction of the two-fold symmetry axis. The two LT structures are mirror images of one another, and their DMs have an angle of roughly 10° with respect to the symmetry axis. The asymmetry can already be seen from the LT FODs: counting all C-C bond FODs in the symmetry-
equivalent parts of the molecule shows that there are 14 FODs on one side, and 16 on the other. This imbalance is then reflected and reinforced when the electron density is optimized in the self-consistent field procedure, as was already mentioned in Section IIIB.

Moreover, as mentioned in Section II A, combining the spin-up FODs of LT1 and the spin-down FODs of LT2 results in a LDQ structure. This LDQ structure has 15 C-C bond FODs in both symmetry-equivalent parts of the molecule, and thereby affords an unbiased description for the molecule. Thus, we can introduce LDQ structures as a useful representation of aromatic bonding. The character of the LDQ structure as a superposition of the LT structures is also observable in the DM: the averaged LT DM is \( \mu_{\text{LT1}+\text{LT2}} = 0.71 \, \text{D} \), it aligns with the molecular axis like the DMs from DFT and LDQ, and has a similar size to the LDQ DM, \( \mu_{\text{LDQ}} = 0.65 \, \text{D} \).

There are several other molecules in the DIP16 set where LT FODs likewise lead to incorrect DMs. In acridine (mol01), LT configurations can lead to DMs with an angle of about 14° with respect to the symmetry axis, while the LDQ DM correctly reflects the symmetry of the molecule. Phenazine (mol12) has LT structures which introduce a significant DM of about 1 D which is oriented perpendicular to the N-N axis, while the DM should be zero according to the molecule’s D\(_{2h}\) point group; only the LDQ solution reproduces the correct DM. Similar results are determined for PyFLOSIC, see the SI, and the observations are analogous. For instance, the stability analysis discussed in Section III A was carried out for the FLO-SIC solutions of azulene (mol03), and they were found to be true local minima.

For either ERKALE or PyFLOSIC, the LDQ DMs are typically in best agreement with experimental values for the absolute DM, and always respect the symmetries of the molecules in this study. However, some LT DMs do not do so, leading to artifactual dipole moments as showcased above for azulene.

2. Non-aromatic molecules

Both LT and LDQ describe identical chemical bonding patterns for non-aromatic parts of the molecules in this study; this means there is only one FOD geometry for such structures. Our test set contains 3 non-aromatic molecules which do not have any aromatic fragments: p-benzoquinone, p-chloranil, and p-fluoranil. These molecules are mol04, mol07, and mol09 in our database; see Table I. They all have a D\(_{2h}\) molecular geometry; thereby, none should have a DM. This is correctly reflected by all methods (DFT, RSIC, CSIC, FLOSIC) used in this study.

For molecules which do have aromatic fragments, like dichlorine (mol08), fluorene (mol10), and 1,4-naphthalenedione (mol11), various electronic geometries can be generated for their aromatic part, allowing for a greater variety in the FOD search space. The resulting DMs point along the C\(_2\) symmetry axis. An exception is the LT3 structure of fluorene (mol10) that has a crooked DM with an angle of about 31° to the symmetry axis in ERKALE and PyFLOSIC. This is unsurprising per the discussion in Section IVA as the corresponding electronic geometry shows 4 C-C bond FODs on one side of the inner C-ring, while there are 6 C-C bond FODs on the other side; this imbalance leads to a bias that tilts the electron density to one side of the molecule during the optimization.

As before, the DMs that arise from various LT and LDQ FOD arrangements turn out to be different in size. DMs from LDQ FODs yield the best agreement of all SIC DMs with experimental reference values.

3. Anti-aromatic molecules

The antiaromatic cyclobutadiene (mol15) molecule has two sets of C-C bond lengths: 1.43 Å and 1.35 Å. According to its D\(_{2h}\) point group, the DM should be zero. Similarly to Section IV A, two LT and one LDQ structure can be formed. Given the distinct C-C bond lengths, one can expect the LT structure in which the double-bond FODs are placed at the shorter C-C bonds to be energetically favorable. Our results are in line with these expectations, that is, the DM is zero in all calculations and the LT structures that place the double-bond FODs in the shorter C-C bonds are lower in energy than the structures that place the double-bond FODs on the longer C-C bonds.

4. Summary

Having discussed the general features of the results on the DIP16 set of molecules, we can proceed with a summary analysis of the DIP16 data. The mean error (ME)

\[
\text{ME} = \frac{1}{N} \sum_i (\mu_{i,\text{calc}} - \mu_{i,\text{ref}}) 
\]

and mean absolute error (MAE)

\[
\text{MAE} = \frac{1}{N} \sum_i |\mu_{i,\text{calc}} - \mu_{i,\text{ref}}| 
\]

are calculated over the entire DIP16 test set, and are analyzed in the following. The explicit values and more details about the used reference values are found in the SI.

The PW92 DFT calculations yield a ME of -0.06 D and a MAE of +0.10 D both in ERKALE and PySCF; the excellent agreement again confirms the reproducibility and comparability of the results. The MAEs for the SIC calculations from various starting points are shown in Table II, the agreement with experiment is worse with SIC than without it, as the errors in Table II are greater...
than the DFT values given above. In Table 3, LT and LDQ stand for initialization with FODs/FLOs arising from these two methods; as some of the molecules afford many possible LT structures, the LT configuration leading to the lowest energy was chosen in each case. An analogous procedure was also employed for LDQ. Finally, $E_{\text{min}}$ refers to using the energetically lowest-lying solution, regardless of whether it represents LT or LDQ type bonding.

Table 3. Mean error (ME) and mean absolute error (MAE) of absolute DMs for DIP16 in D for RSIC calculations in ERKALE and FLO-SIC calculations in PyFLOSIC initialized by various methods (see main text).

| Method       | ME        | MAE       | $E_{\text{min}}$ |
|--------------|-----------|-----------|------------------|
| ERKALE       | LT +0.22  | +0.19     | +0.22            |
|              | LDQ +0.29 | +0.25     | +0.28            |
| PyFLOSIC     | ME +0.19  | +0.01     | +0.21            |
|              | MAE +0.33 | +0.11     | +0.33            |

The RSIC calculations started from FODs in ERKALE result in DMs that agree with the corresponding PyFLOSIC values in most cases. In cases where RSIC and FLO-SIC disagree, we found that the stability analysis and the subsequent re-optimization of the density in ERKALE has taken the RSIC solution away from the FOD starting point; without the stability analysis, the agreement between RSIC and FLO-SIC is recovered. This indicates that FLO-SIC does not correspond to a local minimum of RSIC in such cases.

Interestingly, the best DMs are obtained when starting from LDQ FODs, both in FLO-SIC calculations with PyFLOSIC as well as RSIC calculations ERKALE; in PyFLOSIC, the MAE is close to the DFT value. Worse DMs are obtained when analyzing DMs from FODs based on LT, or the overall energetically lowest solutions ($E_{\text{min}}$) which likewise results in the use of some LT structures instead of LDQ structures.

5. Validation of computational methodology

To emphasize that the results obtained with the PW92 functional and the pc-1 basis set are qualitatively correct, additional DMs of azulene (mol03), see Section IVB1, were computed using PyFLOSIC with other exchange-correlation functionals and larger basis sets. Starting out with the question of the exchange-correlation functional, additional calculations were performed with the PBEsol$^{22}$ and $r^2$SCAN$^{23}$ functionals, which are GGA and meta-GGA approximations, respectively. PBEsol has been identified as a particularly good functional for SIC calculations$^{24}$ whereas $r^2$SCAN exemplifies the state of the art in non-empirical density functional approximations. For these tests, the FODs were re-optimized for each functional. As can be seen from the results in Table 4, the DMs from PBEsol are close to those from PW92, while $r^2$SCAN leads to larger dipole moments than PW92 or PBEsol.

However, most importantly for the present study, the orientation of the DMs are similar in each case: the DMs based on LT FODs are noticeably crooked, confirming that the same issues can be found at other levels of theory and validating the findings of Ref. 15 for the case of FLO-SIC.

Table 4. DMs $\mu$ for azulene (mol03) using various exchange-correlation functionals, as well as the angle $\alpha_{\text{LT}}$ between the DMs from the two LT structures and the molecular symmetry axis, see Fig. 6. The pc-1 basis set was used in all calculations.

| Functional      | $\mu_{\text{DFT}}$ | $\mu_{\text{LT}}$ | $\mu_{\text{LDQ}}$ | $\alpha_{\text{LT}}$ |
|-----------------|---------------------|-------------------|---------------------|----------------------|
| PW92            | 0.96 D              | 0.96 D            | 1.01 D              | 10.1\textdegree      |
| PBEsol          | 0.68 D              | 0.66 D            | 0.97 D              | 9.9\textdegree       |
| $r^2$SCAN       | 0.59 D              | 0.55 D            | 0.86 D              | 6.2\textdegree       |
| $\alpha_{\text{LT}}$ | 10.1\textdegree    | 9.9\textdegree   | 6.2\textdegree     |                      |

Next, the basis question. Repeating the calculations in the split-valence, polarized double-$\zeta$, polarized triple-$\zeta$, and polarized quadruple-$\zeta$ pc-0, pc-1, pc-2, and pc-3 basis sets, respectively, as well as their counterparts augmented with diffuse basis functions, the aug-pc-0, aug-pc-1, aug-pc-2, and aug-pc-3 basis sets, respectively, we obtain the results in Table 5. As these calculations are computationally costly, and as we are only interested in the basis set convergence, the FODs were not reoptimized; instead, they were frozen to the pc-1 optimized values.

Table 5. DMs $\mu$ and the angle $\alpha_{\text{LT}}$ between the LT DMs and the symmetry axis for azulene (mol03) using various pc-$\eta$ and aug-pc-$\eta$ basis sets and the PW92 functional.

| Basis Set       | $\mu_{\text{DFT}}$ | $\mu_{\text{LT}}$ | $\mu_{\text{LDQ}}$ | $\alpha_{\text{LT}}$ |
|-----------------|---------------------|-------------------|---------------------|----------------------|
| pc-0            | 1.13 D              | 0.96 D            | 0.91 D              | 9.2\textdegree       |
| pc-1            | 0.88 D              | 0.68 D            | 0.62 D              | 10.1\textdegree      |
| pc-2            | 0.82 D              | 0.59 D            | 0.52 D              | 10.4\textdegree      |
| pc-3            | 0.78 D              | 0.57 D            | 0.51 D              | 9.9\textdegree       |

The DFT DMs arising from the calculations in the aug-pc-2, pc-3, and aug-pc-3 basis sets are identical, meaning that the DM has converged to the complete basis set limit. Similarly, there is excellent agreement in the SIC DMs between the pc-3 and aug-pc-2 calculations; SIC calculations were not attempted in aug-pc-3 due to the large size of this basis set. We can thus confirm that the data with the augmented triple-$\zeta$ aug-pc-2 basis set is (almost) at the complete basis set limit.

Having established the basis set limit, we can comment on the implications of the basis set convergence on the present study. Although improving the basis affects the absolute values of the DMs, the angle between the LT
DMs to the symmetry axis remains at roughly 10°. This illustrates that the symmetry-broken solutions are not artifacts of small basis sets and the results obtained with the pc-1 basis are reliable.

The sufficiency of the pc-1 basis for the present study is further supported by PW92 DFT calculations for the whole DIP16 set in the aug-pc-2 basis. The ME and MAE values for pc-1 are in excellent agreement with values from the aug-pc-2 calculations, which are −0.05 D and 0.10 D, respectively. Since the DFT dipole moments coincide with the molecular plane, the importance of diffuse functions is less pronounced, as the molecular plane already has a decent amount of flexibility from the valence basis functions in the pc-1 basis set.

As a note, we tested whether fixing the FODs corresponding to the 1s orbitals to the nuclear sites helps to converge the FLO-SIC calculations without compromising the resulting DMs. As can be seen from the results presented in the SI, the DMs resulting from such a procedure are in excellent agreement with the DMs from the fully optimized FODs. Although this strategy was not exploited in this work, it could be used in future investigations to achieve better computational efficiency.

C. Further investigations on bonding motifs

We would like to point out that while we have investigated several types of FOD configurations for the DIP16 database, we have not made any claims to have found the best possible FOD configurations for any of the systems examined in this work. Even though we have examined several kinds of electronic geometries in our calculations, that is, multiple LT and LDQ structures, it is possible that some other type of FOD configuration yields even lower energies. However, all the FOD configurations employed in this work are fully reproducible and can even be generated automatically, thus delivering reproducible FLO-SIC solutions; the molecular and FOD configurations used in this work have also been included in the SI.

For the case of benzene, fascinating bonding motifs in which all three double-bond FOD pairs hover out-of-plane on the same side of the molecule have been found and described by Pederson and Barua.20 For a symmetric molecular geometry for benzene; correspondingly, we will term these hovering FODs. It is likely that several of the molecules in the DIP16 database likewise afford hovering FOD configurations, given that they are cyclic, planar molecules like benzene. However, the role of the hovering bonding motifs is still not clearly understood; at present, we also lack the tools to generate them in an automatic manner. Until now, despite thorough efforts, we have only found stable hovering FODs for ring systems. In contrast, hovering double-bond FOD pairs are not suitable solutions for ethene, (C2H4), for example, as such FODs converge back to a LT solution upon optimization.

![Figure 7. Examples of optimized hovering FODs in benzene.](image)

(a) 1UH  
(b) 2UH  
(c) 3UH  
(d) 3RH

The DIP16 database contains molecules that were also considered by Adhikari et al.25 Ref. 25 describes using fixed molecular geometries from a chemical database, and optimizing only a single FOD configuration for each molecule. Because of the problem of many local minima first pointed out in Ref. 15 for variational SIC calculations and extended to FLO-SIC in this work, it is unlikely that such a procedure leads to the global minimum for FLO-SIC. Comparing our FOD configurations with the ones of [Adhikari et al.25] obtained directly from the authors, we observed that their FOD configurations place some double-bond FOD pairs above the molecular plane for acridine (mol01) and phenazine (mol12); the FOD structures for these molecules are thus partly hovering in nature.

In order to try to understand such solutions, we return to our favorite model system—benzene—and start out by examining the FOD structure described by Pederson and Barua.20 Employing the methodology described in Section II to optimize this configuration, denoted as 3RH in Fig. 7, we obtain a large out-of-plane dipole moment (ca. 0.10 D) and an energy of −233.013171 \(E_h\), which is lower than that of our LT solution (−233.007512 \(E_h\)).

Obviously, such a dipole moment is completely artifactual, and is at variance to experiment and results of ab initio calculations. The artifactual dipole moment can be eliminated for this molecule by splitting the \(\alpha\) and \(\beta\) FODs onto different sides of the molecule; we denote this novel FOD pattern as unrestricted hovering; it is illustrated as 3UH in Fig. 7. However, as the spin symmetry is now broken, this procedure leads to a slightly non-zero spin density and an \(\langle S^2 \rangle \approx 5 \times 10^{-4}\).

However, other kinds of hovering solutions are possible: instead of hovering all three double-bond FOD pairs, one can also hover just one or two of them, and fully relax the FODs in an unrestricted calculation; the structures are presented in Fig. 7. Hovering one pair lowers the energy (1UH, \(E = -233.009569 E_h\)), but induces a large dipole moment (0.28 D) in the molecular plane. Next, hovering two pairs lowers the energy further (2UH, 

\[ h \langle E \rangle \approx 233.007512 E_h \]
The symmetric LDQ solution is higher in energy than the lar geometry for benzene can be captured with SIC by solutions. We demonstrated that a symmetric molecule SIC (CSIC) can be reproduced starting from Lewis-like solutions for benzene in real (RSIC) and complex orbital parts of the manuscript we showed that symmetry-broken that is, the wanted type of local minimum. In the first corrected calculations towards the wanted type of solution, bitals (FLOs) can be used to guide self-interaction corrections (FODs) and the corresponding Fermi–Löwdin or-}


calculations. We have shown that Fermi-orbital describing is an important problem in self-interaction corrected techniques due to the problem of multiple local minima: the two hovering structures employed by Adhikari et al. can likely be further tuned to give a significantly lower energy by hovering more or all of the double-bond FOD pairs.

All FLOs corresponding to hovering FODs appear to be less localized than FLOs from the respective LT or LDQ FODs, both according to visual inspection, as well as according to the Foster–Boys localization criterion. This is another aspect which needs to be studied in more detail in the future. More work is thus needed for a full understanding of any potential hovering FOD solutions. Moreover, it is important to note that each electronic geometry is likely to correspond to a different molecular geometry. When the geometry is optimized for the LT structure, yielding a symmetry-broken molecular geometry with alternating single and double bonds, the hovering FOD solutions no longer yield a decrease in energy.

V. SUMMARY AND DISCUSSION

As originally pointed out in Ref. symmetry breaking is an important problem in self-interaction corrected calculations. We have shown that Fermi-orbital descriptors (FODs) and the corresponding Fermi–Löwdin orbitals (FLOs) can be used to guide self-interaction corrected calculations towards the wanted type of solution, that is, the wanted type of local minimum. In the first part of the manuscript we showed that symmetry-broken solutions for benzene in real (RSIC) and complex orbital SIC (CSIC) can be reproduced starting from Lewis-like solutions. We demonstrated that a symmetric molecular geometry for benzene can be captured with SIC by starting the calculation from a Linnett double-quartet (LDQ) structure in RSIC, CSIC, and in FLO-SIC. While the symmetric LDQ solution is higher in energy than the Lewis structures when the molecular geometry is allowed to relax, the LDQ structure is the lowest-lying solution for the symmetric molecular geometries considered in this study. Moreover, the millihartree energy difference between Lewis’ theory (LT) and LDQ structures is negligi-}

le with respect to the overall SIC contribution which is in the order of hartrees.

In the second part of this study, a benchmark set consisting of 16 planar, cyclic molecules we call DIP16 was used to show that Lewis-like SIC solutions often break the symmetry of the electronic density, while LDQ SIC solutions do not. Reasonable LDQ solutions can typically be constructed by combining two Lewis-like solutions, by taking the spin-up FODs from one Lewis structure and the spin-down FODs from another Lewis structure. Thus, we were able to demonstrate that LDQ SIC is able to provide a simple representation of resonance/ar-}


![Figure 8. DIP16: Comparison of LDQ DMs and the corresponding averages of LT DMs from PyFLOSIC calculations; the DMs as well as the averaging procedure can be found in the SI. See Table 1 for the legend of molecules.](image-url)
of the molecule differ, the DM from a self-consistent calculation will point towards the part of the molecule that has more FODs; that is, the density follows the FOD arrangement. Any imbalances in the FOD arrangement result in a bias for the electron density in SIC calculations, as the SIC prefers localized orbitals that are determined by the FODs.

Our findings are not limited to the FLO-SIC method, as we were also able to reproduce the main findings of this work also for RSIC and CSIC by starting the calculations from FLOs guided by LT or LDQ FODs.

A careful initialization of SIC calculations is of paramount importance. For instance, various starting points may lead to convergence to different local minima. In many cases, several FOD arrangements have to be evaluated to find the minimum-energy SIC solution. Employing a single type of FOD arrangement may lead to incorrect conclusions.

Concluding the summary, we move on to further discussion. The second part of this study—the examination of the DIP16 dataset described in this work—employed fixed molecular geometries. Optimizing the geometries will likely bias the dipole moments further if the FODs break the molecular symmetry. New advances in theory are warranted in order to improve the usefulness of self-interaction corrected calculations for practical applications, which typically require the determination of optimal molecular geometries, for instance, and we hope to address such topics in future work.

Furthermore, we excluded hovering FOD structures from the current investigations on the DIP16 dataset. Although all presently-known hovering structures can be constructed starting from a LT or LDQ configuration by moving some of the FODs out-of-plane, there is currently no elegant, systematic way to set up reproducible hovering motifs for more complex molecules, such as those included in the DIP16 benchmark set of this work. The phenomenon of hovering FODs is not yet understood, and we hope to find suitable explanations in further research.

As shown in this study, several kinds of FOD configurations and FLO-SIC solutions are possible for molecules as simple as benzene, as could be expected from the results of Ref. [15] for fully variational SIC. We have demonstrated in Section 11 that the various kinds of FLO-SIC solutions all correspond to local minima at symmetric nuclear geometries for benzene. Importantly, as shown here and in Ref. [15] if the nuclear geometry is relaxed, the resulting optimal molecular geometries may be different for each type of solution, further complicating the problem of locating the lowest-lying solution.

Finally, we would also like to point out that free and open-source software is essential for a fruitful development of computational methods, as they enable straightforward one-to-one comparisons as in this work. All authors of this paper are either main developers of, or contributors to various free and open-source software electronic structure programs, making self-interaction correction methods readily available to a broad community of physicists, chemists, and material scientists.

ACKNOWLEDGMENT

We thank all the sponsors and guests of the "Quo Vadis Self-Interaction Correction?" workshop (Freiberg, Germany, 2019), as this meeting brought some of the present authors together in person for the first time and thereby enabled them to start a fruitful cooperation. Moreover, it was an honor to meet Prof. John P. Perdew, the heart and one of the most important persons in SIC and a living legend in the field of electronic structure. Further thanks goes to Prof. Mark R. Pederson for creating the FLO-SIC method and Prof. Hannes Jónsson for major contributions in the development of the CSIC method, which are both methods discussed within the manuscript. This contribution is part of our OpenSIC project, which aims to develop fresh perspectives in the field of self-interaction corrections using open-source software. J. Kortus and S. Schwalbe have been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project ID 421663657 - KO 1924/9-1. We acknowledge the contributions of Jakob Kraus at an early stage of this project. We furthermore thank the ZIH in Dresden for computational time and support.

SUPPORTING INFORMATION

See the supporting information for thorough information on the used molecular geometries, the full data set of dipole moments, comparison of the LDQ dipole moments as averages of the respective LT dipole moments, the molecule-by-molecule discussion of the electronic geometries and the resulting dipole moments of the DIP16 data set, as well as the analysis of freezing the 1s FODs at the nuclei. The full set of molecular and electronic geometries is also available as part of the supporting information.

DATA AVAILABILITY STATEMENTS

The data that supports the findings of this study are available within the article and its supporting information. The data are also openly available in [https://gitlab.com/kaitrepte/dip16]

AUTHOR DECLARATIONS

The authors have no conflicts to disclose.
Phys. Rev. B 85, 052505 (2017).

P.-O. Löwdin, “On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals,” J. Chem. Phys. 18, 365 (1950).

X. Schwolbe, L. Fiedler, J. Kraus, J. Kortus, K. Trepte, and S. Lehtola, “PyFLOSIC: Python-based Fermi-Löwdin orbital self-interaction correction,” J. Chem. Phys. 153, 084104 (2020).

D. Halt and M. Head-Gordon, “How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values,” J. Chem. Theory Comput. 14, 1969–1981 (2018).

A. Adhikari, B. Santra, S. Ruan, P. Bhattacharj, N. K. Nepal, K. A. Jackson, and A. Ruzsinszky, “The Fermi-Löwdin self-interaction correction for ionization energies of organic molecules,” J. Chem. Phys. 153, 184303 (2020).

26R. D. Johnson, NIST Computational Chemistry Comparison and Benchmark Database (NIST Standard Reference Database Number 101, August–21 (2020).

27K. C. Hafner, A. M. Martin, N. Patel, M. Shevchuk, N. Kau, A. Tran, J. Tseityn, D. S. Graham, Y. Niyonzima, S. E. Wollman, A. M. Newman, A. M. Zhang, S. F. Dermer, E. N. Ho, S. Aggarwal, E. W. Jin, S. Pan, M. Y. Liou, J. K. Skerritt, H. M. Park, N. B. Ravi, S. C. Ness, D. X. Du, J. W. Qi, A. H. Yang, T. C. Allison, K. K. Irikura, and J. F. Liebman, Computed 3-D Structures (NIST Chemistry WebBook, SRD 69, February–9 (2015) note: 3D structures are B3LYP/6-31G* quality.

28http://www.chemspider.com/Erkale.htm (2021).

29S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, “PubChem in 2021: new data content and improved web interfaces,” Nucleic Acids Res. 49, D1388–D1395 (2020).

30I. Lehtola, M. Hakala, A. Sakko, and K. Hämäläinen, “ERKALE - A flexible program package for X-ray properties of atoms and molecules.” J. Comput. Chem. 33, 1572 (2012).

31S. Lehtola. ERKALE- HF/DFT from Hel, http://github.com/susilehtola/erkale/ (2016).

32Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Erikson, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamsom, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, “Recent developments in the PySCF program package,” J. Chem. Phys. 153, 024109 (2020).

33S. Lehtola and A. Karttunen, “Free and open source software for computational chemistry education,” (2021).

10.33774/chemrxiv-2021-hr1r0.

34S. Lehtola, C. Steigemann, M. J. T. Oliveira, and M. A. L. Marques, “Recent developments in LIBXC – A comprehensive library of functionals for density functional theory,” SoftwareX 7, 1 (2018).

35J. P. Perdew and Y. Wang, “Accurate and simple analytic representation of the electron-gas correlation energy,” Phys. Rev. B 45, 13244–13249 (1992).

36F. Bloch, “Bemerkung zur Elektronentheorie des Ferromagnetismus und der elektrischen Leitfähigkeit,” Z. Phys. 57, 545–559 (1929).

37P. A. M. Dirac, “Note on Exchange Phenomena in the Thomas Atom,” Math. Proc. Cambridge Philos. Soc. 26, 376–385 (1930).

38P. Jensen, “Polarization-consistent basis sets: Principles,” J. Chem. Phys. 115, 9113 (2001).

39P. Jensen, “Polarization consistent basis sets. II. Estimating the Kohn–Sham basis set limit,” J. Chem. Phys. 116, 7372 (2002).

40J. Lehtola, P. Manninen, M. Hakala, and K. Hämäläinen, “Complements-optimized basis sets: Application to ground-state electron momentum densities.” J. Chem. Phys. 137, 104105 (2012).

41P. Porezag and M. R. Pederson, “Optimization of Gaussian basis sets for density-functional calculations,” Phys. Rev. A 60, 2840 (1999).

42O. Treutler and R. Ahrlich, “Efficient molecular numerical integration schemes,” J. Chem. Phys. 102, 346 (1995).

43A. D. Becke, “A multicenter numerical integration scheme for polatomic molecules,” J. Chem. Phys. 88, 2547–2553 (1988).

44S. Lehtola and H. Jónsson, “Unitary optimization of localized molecular orbitals,” J. Chem. Theory Comput. 9, 5305–5372 (2013).

45A. L. McClellan, Tables of experimental dipole moments, Vol. 3 (WH Freeman, 1963).
A. Gelessus, W. Thiel, and W. Weber, “Multipoles and Symmetry,” J. Chem. Educ. 72, 505 (1995).

F. L. Tobaison, L. Huestis, C. Chandler, S. E. Pedersen, and P. Peters, “The polar nature of 2,1,3-benzoxadiazole, -benzothiadiazole, -benzoselenadiazole and derivatives as determined by their electric dipole moments,” J. Heterocycl. Chem. 10, 773–778 (1973).

T. Satoh, M. Kudoh, T. Tsuji, S. Kita, T. Morii, and S. Sudoh, “Theoretical and experimental studies on the ground- and excited-state dipole moments of 1,4-naphthoquinone and its derivatives,” Bull. Chem. Soc. Jpn. 80, 1103–1113 (2007).

Y. S. El-Sayed, “Spectrophotometric and fluorescence quenching studies of the charge transfer complexation of (4-dimethylamino- benzylidene)-(4,6-dimethyl-pyrimidin-2-yl)-amine with some organic acceptors,” Spectrochim. Acta, Part A 78, 1227–1233 (2011).

V. Dobyns and L. Pierce, “Microwave spectrum, structure, dipole moment, and quadrupole coupling constants of 1,2,5-thiadiazole,” J. Am. Chem. Soc. 85, 3553–3556 (1963).

M. R. Battaglia, A. D. Buckingham, and J. H. Williams, “The electric quadrupole moments of benzene and hexafluorobenzene,” Chem. Phys. Lett. 78, 421–423 (1981).

J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, “Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces,” Phys. Rev. Lett. 100, 136406 (2008).

J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, “Accurate and Numerically Efficient r^2SCAN Meta-Generalized Gradient Approximation,” J. Phys. Chem. Lett. 11, 8208–8215 (2020).

S. Lehtola, E. Ö. Jónsson, and H. Jónsson, “Effect of Complex-Valued Optimal Orbitals on Atomization Energies with the Perdew–Zunger Self-Interaction Correction to Density Functional Theory,” J. Chem. Theory Comput. 12, 4296–4302 (2016).

S. Adhikari, private communication (2021).

M. J. T. Oliveira, N. Papier, Y. Pouillon, V. Blum, E. Artacho, D. Caliste, F. Corsetti, S. de Gironcoli, A. M. Elena, A. García, V. M. García-Suárez, L. Genovese, W. P. Huhn, G. Huhs, S. Kokott, E. Küçükbenli, A. H. Larsen, A. Lazzaro, I. V. Lebedeva, Y. Li, D. López-Durán, P. López-Tarifa, M. Lüders, M. A. L. Marques, J. Minar, S. Mohr, A. A. Mostofi, A. O'Cais, M. C. Payne, T. Ruh, D. G. A. Smith, J. M. Soler, D. A. Strubbe, N. Tancogne-Dejean, D. Tildesley, M. Torrent, and V. W. Yu, “The CECAM electronic structure library and the modular software development paradigm,” J. Chem. Phys. 153, 024117 (2020).