

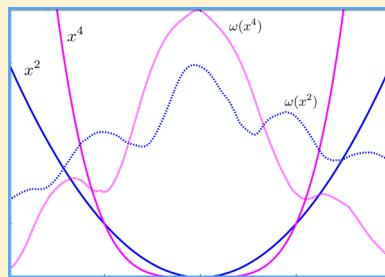
Characterization and Generation of Local Occupied and Virtual Hartree–Fock Orbitals

Ida-Marie Høyvik^{*,†} and Poul Jørgensen[‡]

[†]Department of Chemistry, The Norwegian University of Science and Technology, Høgskoleringen 5, 7491 Trondheim, Norway

[‡]qLEAP Center for Theoretical Chemistry, Department of Chemistry, Aarhus University, Langelandsgade 140, 8000 Aarhus C, Denmark

ABSTRACT: The scope of this review article is to discuss the locality of occupied and virtual orthogonal Hartree–Fock orbitals generated by localization function optimization. Locality is discussed from the stand that an orbital is local if it is confined to a small region in space. Focusing on locality measures that reflects the spatial extent of the bulk of an orbital and the thickness of orbital tails, we discuss, with numerical illustrations, how the locality may be reported for individual orbitals as well as for sets of orbitals. Traditional and more recent orbital localization functions are reviewed, and the locality measures are used to compare the locality of the orbitals generated by the different localization functions, both for occupied and virtual orbitals. Numerical illustrations are given also for large molecular systems and for cases where diffuse functions are included in the atomic orbital basis. In addition, we have included a discussion on the physical and mathematical limitations on orbital locality.



CONTENTS

1. Introduction	3306
Localization Functions	3307
Spatial Locality of Orbitals and Sets of Orbitals	3307
Localizability of Hartree–Fock Orbitals	3307
Localization Function Optimization	3307
Generation of Localized Orbitals: Top-Down and the Bottom-Up Localization Strategy	3308
Starting Orbitals Using a Top-Down Orbital Localization Strategy	3308
Scope and Outline	3309
2. Characterizing Orbital Locality	3309
2.1. Second Moment Orbital Spread	3309
2.2. Fourth Moment Orbital Spread	3309
2.3. Tail Spread	3309
2.4. Reference Point: Atomic Orbitals	3310
2.5. Reference Point: Projected Atomic Orbitals	3310
3. Orbital Localization Functions	3310
3.1. Pipek–Mezey Localization Function	3310
3.2. Edmiston–Ruedenberg Localization Function	3311
3.3. Boys Localization Function	3312
3.4. Powers of the Second Central Moment	3312
3.5. Powers of the Fourth Central Moment	3312
4. Exemplifying Orbital Locality Measures	3313
4.1. Locality of Individual Orbitals	3313
4.2. Locality of a Set of Orbitals	3314
5. Constraints on the Localizability of a Set of Molecular Orbitals	3314
5.1. Orbital Bases for Separating the Physical and Mathematical Constraints on Locality	3314

5.2. Numerical Illustrations for Separating the Physical and Mathematical Effects	3315
6. Locality Illustrations for Single Orbitals	3316
6.1. Occupied Orbitals	3317
6.2. Virtual Orbitals	3317
7. Locality of a Set of Orbitals	3318
7.1. Locality of the Occupied Orbitals	3318
7.2. Localizing the Virtual Orbitals for Non-augmented Basis Set	3319
7.3. Local Virtual Orbitals from Basis Sets Augmented with Diffuse Functions	3320
8. Exploring Orbital Locality for Local Correlation Methods	3321
9. Concluding Remarks	3323
Author Information	3324
Corresponding Author	3324
Notes	3324
Biographies	3324
Acknowledgments	3324
References	3324

1. INTRODUCTION

Local occupied molecular orbitals (MOs) were introduced into quantum chemistry in 1949 by Lennard-Jones, Pople, and Hall,^{1–5} and since then, a vast number of methods to generate local orbitals have been developed.^{6–40} The motivation for generating local orbitals is both conceptual and computational. The occupied molecular orbitals describe the electron density,

Received: August 24, 2015

Published: February 8, 2016

and the locality of the occupied orbitals therefore is intimately connected with the electronic structure of a molecular system. From a conceptual point of view, local occupied orbitals therefore offer the possibility to describe and analyze molecular structures and bondings in terms of chemically intuitive local orbitals.^{41–49} The local virtual orbitals offer less physical insight with respect to the description of the electronic structure of a molecular system. However, a local virtual orbital space is essential for developing so-called local correlation methods where the local nature of electron correlation effects is exploited by expressing the correlated wave function in a basis of local orbitals. Local correlation methods date back to the 1960s,^{50–52} and they have grown increasingly popular. In local correlation methods, the occupied Hartree–Fock orbital space is spanned by a set of localized Hartree–Fock orbitals and the virtual Hartree–Fock orbital space is usually spanned by a set of projected atomic orbitals (PAOs),^{53–56} pair natural orbitals,^{57–59} or orbital-specific virtuals.^{60,61} More recently, the virtual Hartree–Fock orbital space has also been spanned by a set of localized virtual Hartree–Fock orbitals.^{62,63} Local correlation methods have predominantly been developed for the treatment of dynamical correlation^{53–55,58,60,62–81} (in particular for coupled cluster methods) but have also found some use for the treatment of static correlation using multireference approaches.^{82–86} An overview of local correlation methods is given by Beran and Hirata,⁸⁷ and an overview of the developments in local coupled cluster methods for properties is given by Crawford.⁸⁸

Localization Functions

Local orbitals may be generated by exploiting the invariance of the Hartree–Fock wave function with respect to orthogonal transformations among the occupied and among the virtual orbitals. The most popular and successful approaches for orbital localization are based on iterative schemes where orthogonal transformations are used to optimize an orbital localization function. Three orbital localization functions have been exceedingly popular and widely used for localizing the occupied orbital space. These are the localization functions of Boys (as modified by Edmiston and Ruedenberg),^{6–8} Edmiston and Ruedenberg,^{8,9} and Pipek and Mezey.¹⁵ The Boys localization minimizes the sum of the orbitals' second central moments, the Pipek–Mezey localization scheme minimizes the number of atomic centers over which an orbital extends, and the Edmiston-Rudenberg localization function maximizes the sum of orbital self-repulsion energies. Recently, more advanced localization functions have been developed, where the focus has been on increasing the bulk locality of the least local orbitals and on reducing the thickness of the orbital tails. Jansík et al.³² imposed a penalty on orbitals with a large second central moment by introducing powers in the Boys localization function (powers of the second central moments) to improve bulk locality. Høyvik et al.³⁴ suggested to use powers of the orbitals' fourth central moments to reduce the thickness of orbital tails. Powers of the Pipek-Mezey localization function were reported by Lehtola et al.,⁸⁹ but in contrast to the case for the Boys localization function, powers of the Pipek-Mezey localization function were not seen to give locality improvements.

Spatial Locality of Orbitals and Sets of Orbitals

The spatial locality of orbitals is critical for the efficiency of local correlation methods. This raises the question how to measure if an orbital or a set of orbitals is local. In the early

years of orbital localization, an orbital or a set of orbitals were classified as local if the orbitals were obtained from a minimization of a localization function. However, more objective measures which more directly reflects the spatial extent of an orbital have to be used. Many different measures have been introduced, ranging from general measures such as the second^{14,24,30,32,90,91} and fourth³⁴ moment orbital spreads, charge populations,^{18,92,93} and MO coefficient sparsity,^{12,29,31} to more specific measures directly related to a particular localization scheme³⁶ and to a purely visual presentation.^{89,94} We will discuss orbital locality in a context where an orbital is local if it is confined to a small volume in space. Hence, we focus on measures of locality that directly reflect the spatial extent of an orbital. To obtain a complete picture of the spatial extent of an orbital, knowledge is required both of the confinement of the bulk of the orbital and of the thickness of the orbital tails. Inspired by probability theory and statistics where second and fourth central moments are used for describing distributions,⁹⁵ we will use the second and fourth moment orbital spreads as measures for characterizing the bulk and tail of an orbital. Further, for a lucid discussion on whether an orbital with a certain spatial extent is local or not, a reference point is needed. A natural reference point for the locality of MOs, are the AOs from which the MOs are constructed. To characterize the locality of a set of occupied or virtual Hartree–Fock orbitals, we will discuss the use of average and maximum orbital spreads to distinguish sets where orbitals on average are local from sets in which all orbitals in the set are local. The locality characterization tools we use are completely general and may be applied to any set of orbitals. This enables a comparison of orbital locality for orbitals generated by different localization functions, as well as a comparison with AOs and PAOs.

Localizability of Hartree–Fock Orbitals

When discussing the locality of orthogonal Hartree–Fock orbitals, one must also consider the limitations that may exist on the orbitals' localizability. Since the occupied orbitals must be able to describe the electron density, the localizability will be dependent on the locality of the electron density as discussed by e.g., Millie et al.,⁹⁶ Daudel et al.,⁹⁷ and Malrieu et al.⁹⁸ In general terms, the localizability of a set of occupied and virtual orbitals will depend on the physical constraint that is imposed on the orbital basis by the Hartree–Fock optimization condition (i.e., the separation into an occupied and a virtual orbital space). Further, the localizability is limited by the mathematical constraint that is imposed by requiring that the orbitals are orthogonal. The implications of the two constraints on the orbital locality has been discussed by Høyvik et al.⁹⁹

Localization Function Optimization

To optimize localization functions, Edmiston and Ruedenberg proposed in 1963⁸ to perform a sequence of pairwise rotations. This approach, the Jacobi sweeps of iterations, has since become the standard approach for optimizing localization functions. The Jacobi sweeps work well for the occupied orbital space where convergence may be obtained in a relatively straightforward manner. The success of the Jacobi sweeps for the occupied orbital space may be connected to the fact that most localization functions have simple hypersurfaces with well isolated and strong minima for the occupied orbital space.²⁴ In contrast, Jacobi sweeps of iterations have failed to generate local virtual orbitals for all but simple molecules in small bases.

Since 1963, a large number of optimization algorithms and schemes have been suggested for optimizing localization functions. In a broad sense, these can be divided into two classes: (i) first-order methods which only require that the gradient of the localization function can be calculated, and (ii) second-order methods where information about both the gradient and the Hessian is used. Minimizations where gradient information is used may be expected to perform better than the Jacobi sweeps, and numerous gradient based methods have been suggested.^{100–103} However, none of these methods have been found to represent significant improvements over Jacobi sweeps and they have failed when applied to localize virtual orbitals. The major problem when applying first-order methods for localizing virtual orbitals is that the localization functions, for all starting guesses of orbitals that are available, have a large number of negative Hessian eigenvalues, and the gradient therefore cannot correctly predict in which direction to move when a minimization is carried out. Leonard et al.¹⁰⁴ proposed to use a Newton–Raphson-like method, either alone or in combination with a first-order method. The Newton–Rapshon method is quadratically convergent, but if the Hessian for the starting guess of orbitals is not positive-definite, this approach will not converge. Quasi-Newton methods such as the Broyden–Fletcher–Goldfarb–Shanno algorithm have also been used for orbital localization,^{31,105} but these methods suffer from the same problems as the Newton method. A general conclusion to the aforementioned optimization algorithms is that they are able to localize the orbitals of the occupied space but fail to localize the orbitals of the virtual space due to the large negative Hessian eigenvalues that occur in the initial iterations of an orbital localization.

Recently, it was shown that the optimization of localization functions for both occupied and virtual Hartree–Fock orbitals may be performed in a straightforward manner by using the trust-region method.¹⁰⁶ The basic idea in the trust-region method is to determine a second-order Taylor expansion of the localization function and identify a region, the so-called trust region, where the second-order Taylor expansion approximates well the localization function. The minimum of the localization function is then determined by performing a sequence of iterations. In each iteration, the minimum of the trust region is identified by solving a set of level-shifted Newton equations. For each iteration, the step to the minimum at the boundary of the trust-region ensures that we are moving downhill. When solving the level-shifted Newton equations, explicit computation and storage of the Hessian is avoided by solving the equations using iterative algorithms,¹⁰⁷ where only linear transformations of the Hessian on trial vectors are required. The trust-region method was introduced by Fletcher¹⁰⁶ and was adapted by Høyvik and Jørgensen to include a line search step to make it an efficient and reliable algorithm for optimizing localization functions for both the occupied and the virtual orbitals, for large molecular systems containing diffuse basis functions,¹⁰⁸ and also when large negative Hessian eigenvalues are encountered in the initial iterations.¹⁰⁹

Generation of Localized Orbitals: Top-Down and the Bottom-Up Localization Strategy

The most common approach for obtaining a set of localized orbitals is to start out performing a Hartree–Fock optimization to obtain a set of occupied and virtual Hartree–Fock orbitals and then perform a localization of these orbitals using one of the localization functions discussed previously. Such a strategy

for generating a set of local orbitals is denoted a top-down strategy,^{37–39} emphasizing that the Hartree–Fock optimization step (the top step) is followed by an orbital localization step (the down step). The generation of local orbitals using the top-down strategy will be discussed later.

The top-down strategy should be contrasted to the bottom-up strategy where a local basis is initially generated, that span the basis of the full molecular calculation, and contains a subset of basis functions that approximate well the occupied set of Hartree–Fock orbitals. For this local basis, the smallest transformation matrix that transform the orbitals from the local basis to a set of Hartree–Fock orbitals is determined. The smallest transformation matrix may be obtained using the least-change algorithm.¹¹⁰ In the bottom-up strategy, the local set of Hartree–Fock orbitals is obtained without recourse to orbital localization functions.

The bottom-up strategy has been used by Liu and co-workers,^{37,38} where the initial local basis is generated by using a fragment in a molecule approach. In this approach, a fragmentation of the molecule is performed and the initial local basis is generated from the orbitals of the individual molecular fragments. Ziolkowski et al.³⁰ have also used the bottom-up approach, where the least-change algorithm¹¹⁰ is used both at the valence and the full molecular basis level. Zhang and Li have also used the least-change approach in their sequential transformation strategy to obtain regionally localized molecular orbitals.³⁹

The locality of orbitals generated from bottom-up approaches is seen to be similar to that obtained using the traditional Boys and Pipek–Mezey localization functions.^{30,37–39} In this review, we restrict ourselves to consider only how local orbitals may be generated when orbital localization functions (top-down strategies) are used.

Starting Orbitals Using a Top-Down Orbital Localization Strategy

When a top-down localization strategy is used, a set of starting orbitals is needed. Canonical molecular orbitals (CMOs) obtained from diagonalizing the Fock matrix may be used. However, the CMOs are completely delocalized across the molecular system, and even though the trust-region method will converge starting from such orbitals, it is computationally more efficient to start out with a set orbitals that have some spatial locality. Several schemes which may generate a set of semilocal orbitals (as compared to the locality of the AOs) without optimizing a localization function have been proposed. Schemes for generating starting orbitals should be computationally inexpensive and generate starting orbitals of good locality. Aquilante et al.²⁸ exploited the locality of the density matrix to generate local orbitals by Cholesky decomposing the density. Cholesky orbitals are computed in a noniterative manner, but it relies on the ordering of the AOs. The Cholesky orbitals were used by Guo et al.¹¹¹ to generate Boys localized orbitals. Høyvik et al. used the bottom-up strategy of Ziolkowski et al. with an explicit localization of the valence basis to generate sets of local starting orbitals.¹¹² In this way, the occupied starting orbitals are well-localized and similar to the ones of the local full molecular basis orbitals. Liu and co-workers have also shown how their bottom-up orbitals may be used as starting orbitals for further localization.^{37,38}

For local correlation and frozen core calculations, the core and valence space localizations should be performed separately to avoid a mixing of the core and valence orbitals. Høyvik et al.

have generalized the approach of Ziolkowski et al. to include a separation of the core and valence orbital spaces.¹¹² It should be noted that the core and valence space have to be treated separately if a least-change approach is used, otherwise, an extensive mixing of the core and valence spaces will occur.¹¹²

Scope and Outline

The scope of this review is to discuss the locality of occupied and virtual Hartree–Fock orbitals generated by both traditional and more recent localization functions. Locality will be discussed from the stand that an orbital is local if it is confined to a small region in space. We have organized the article as follows. In section 2, we discuss how to characterize orbital locality, by focusing on locality measures that directly reflect the spatial extent of an orbital, such as second and fourth moment orbital spread measures. In section 3, the traditional orbital localization functions (the Boys, Pipek–Mezey, and Edmiston–Ruedenberg localization functions) are described, together with more recent developments such as powers of the second and fourth central moments. In section 4, we provide numerical examples of the locality measures and discuss how to report the locality for individual orbitals and for sets of orbitals. In section 5, we discuss the limitation that is imposed on the locality by the Hartree–Fock optimization condition and the constraints imposed by having a set of orthogonal orbitals. Equipped with locality measures and reference points, we discuss in section 6 the locality that is obtained for individual orbitals. In particular, we compare bulk size and tail thickness for orbitals that are localized using various localization functions. In section 7, the locality for orbital sets is discussed. In section 8 the locality of charge density distributions is discussed, and finally, in section 9, we give a summary of the topics treated in this article.

2. CHARACTERIZING ORBITAL LOCALITY

We will now discuss how to characterize the spatial locality of an orbital. To obtain a complete picture of the spatial locality, the spatial confinement of the bulk of the orbital and the orbital tails must be known. We therefore discuss measures to characterize both the size of the bulk and the tail. Further, as a reference point for orbital locality, we use the locality of the AOs since they are the building blocks of the MOs.

2.1. Second Moment Orbital Spread

A widely used orbital locality measure is the square root of the orbital's second central moment,^{14,24,30,32,90,91,113} commonly referred to as the orbital spread. The second central moment (the variance) of an orbital $\phi_p |p\rangle$, in one dimension is defined as,

$$\mu_2^p = \langle p|(\hat{x} - \langle p|\hat{x}|p\rangle)^2|p\rangle \quad (1)$$

where $\langle p|p\rangle = 1$, and $\bar{x}_p = \langle p|\hat{x}|p\rangle$ is the average position of the orbital. The second central moment is thus defined in terms of an integration over the orbital density distribution function

$$\omega_p(x) = \phi_p(x)\phi_p^*(x) \quad (2)$$

and the second moment weight function $(x - \bar{x}_p)^2$. A local orbital has a small second central moment, since the bulk of the density distribution function is located close to its average position. The orbital spread, σ_2^p , given by

$$\sigma_2^p = \sqrt{\mu_2^p} \quad (3)$$

is a measure which characterizes how much the orbital density on average deviates from the orbital's mean position. We term the orbital spread in eq 3 the second moment orbital spread.

It is useful to connect second moment orbital spreads to normal distribution functions, which for orbital $\phi_p(x)$ with the second moment orbital spread σ_2^p , is given as

$$\mathcal{N}(x, \sigma_2^p) = \frac{1}{\sigma_2^p \sqrt{2\pi}} \exp\left[-\frac{(x - \bar{x})^2}{2(\sigma_2^p)^2}\right] \quad (4)$$

The normal distributions generated from a second moment spread may be used as a tool to visualize bulk extents (this will be done in section 6).

2.2. Fourth Moment Orbital Spread

The second moment orbital spread (eq 3) describes where the bulk of the orbital is located. However, a small second moment orbital spread may be obtained because of an acute peak and a thick tail or because of a broad peak and a thin tail. To introduce a measure which puts more weight on the tail region than does the second moment orbital spread, we consider the fourth central moment,⁹⁵

$$\mu_4^p = \langle p|(\hat{x} - \langle p|\hat{x}|p\rangle)^4|p\rangle \quad (5)$$

which contains an integration over the orbital distribution function and a fourth moment weight function. In Figure 1, the

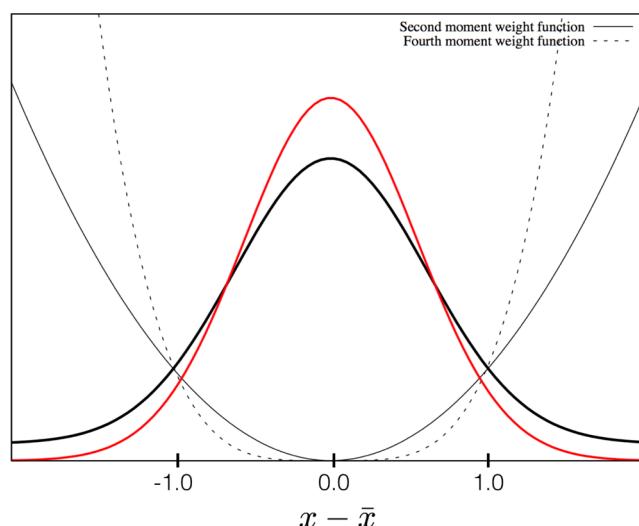


Figure 1. Illustrations of distribution functions where the distribution with $\beta_p \approx 1$ (red line) shows a faster tail decay than the distribution with $\beta_p > 1$ (thick, black line). The second (thin, black line) and fourth moment (thin, dashed line) weight functions are also plotted, to show the increased weight that is imposed on the tail region by the fourth moment weight function compared to the second moment weight function.

relationship between the second and fourth moment weight functions, $(x - \bar{x}_p)^2$ (full bold line) and $(x - \bar{x}_p)^4$ (dashed line), is illustrated. From Figure 1, it is clear that the fourth moment weight function is smaller than the second moment weight function in the region $|x - \bar{x}_p| < 1$, while it is larger than the second moment weight function in the region $|x - \bar{x}_p| > 1$. The difference between the two weight functions further increases when the size of $|x - \bar{x}_p|$ increases. An integration where the orbital distribution function is multiplied with a fourth moment weight function (the fourth central moment, eq 5) and will

therefore put more weight on the tail region than an integration that uses a second moment weight function. If the second central moment is viewed as targeting the bulk of the orbital, the difference between the fourth and second central moment may be viewed as targeting the tail region. Hence, we introduce the fourth moment orbital spread

$$\sigma_4^p = \sqrt[4]{\mu_4^p} \quad (6)$$

as a complementary measure to the second moment orbital spread σ_2^p , where more emphasis is put on the tail decay.

2.3. Tail Spread

In statistics, the kurtosis of a distribution is given by

$$\beta_p^4 = \frac{(\sigma_4^p)^4}{(\sigma_2^p)^4} \quad (7)$$

and it is used as a measure of the tail thickness of distributions,^{114,115} since a larger kurtosis means more of the second central moment is a result of few large deviations, in contrast to many small deviations. For orbital locality, we introduce the tail spread, defined as the fourth root of the kurtosis

$$\beta_p = \frac{\sigma_4^p}{\sigma_2^p} \quad (8)$$

as a measure of tail thickness. If $\beta_p \approx 1$, the fourth moment weighting of the orbital distribution function exhibits little difference compared to a second moment weighting. This indicates that the orbital tails are small, since the second and fourth moment weight functions differ most in the tail region (see Figure 1). Contrary, if $\beta_p > 1$, a significant amount of orbital distribution function is in the tail region. Illustrations of distribution functions with $\beta_p \approx 1$ (red line) and $\beta_p > 1$ (thick, black line) are also plotted in Figure 1. The red curve quickly decays toward zero away from its central position, while the black curve exhibits thick tails. The second and fourth moment weight functions are also displayed in Figure 1 to illustrate the increased weight put on tail regions by the fourth moment weight function.

2.4. Reference Point: Atomic Orbitals

To judge whether an occupied or virtual Hartree–Fock orbital is local, its locality should be compared to the locality of the atomic basis functions that are used for constructing the orbital. To illustrate the locality of the atomic basis functions, we have in Table 1 reported the orbital spreads, σ_2 and σ_4 , and tail spread, β , for the least local atomic basis functions of Dunning's augmented and nonaugmented bases for cardinal numbers D, T, and Q^{116–118} for the carbon and sulfur atoms. For an illustration, we have in Figures 2 and 3 plotted the least local atomic basis function for the carbon atom in the cc-pVTZ and

Table 1. σ_2 (au), σ_4 (au), and β for the Least Local AOs of Carbon and Sulfur for Different Basis Sets

basis	carbon			sulfur		
	σ_2	σ_4	β	σ_2	σ_4	β
cc-pVDZ	2.9	3.1	1.07	3.0	3.2	1.07
cc-pVTZ	3.2	3.5	1.09	3.4	3.7	1.09
cc-pVQZ	3.5	3.8	1.09	3.7	4.1	1.11
aug-cc-pVDZ	5.6	6.1	1.09	5.6	6.1	1.09
aug-cc-pVTZ	5.9	6.4	1.08	6.0	6.5	1.08

aug-cc-pVTZ bases, respectively. From Table 1, we see that the least local AOs for carbon and sulfur have similar orbital spread values. The orbital spreads are between 3.0 and 4.0 au for nonaugmented basis and between 5.5 and 6.5 au for the augmented basis, increasing slightly with an increasing cardinal number. The most diffuse atomic basis functions consist of primitive Gaussian functions and have vanishing tails as measured in terms of the tail spread.

To obtain the density distribution function of the AO in one dimension, we have integrated out the x - and y -components of the orbital density to get the distribution along the z -axis

$$\omega_p(z) = \int \int \phi_p(x, y, z) \phi_p(x, y, z) dx dy \quad (9)$$

In Figure 4, we have plotted the density distribution function along the z -axis for the p-functions plotted in Figures 2 and 3. The figure illustrates the large spatial extent (large σ_2) of both AOs, but in particular, the one for the aug-cc-pVTZ basis, and the rapid tail decay beyond the bulk (small β).

When discussing the sizes of σ_2^p , σ_4^p , and β_p for a Hartree–Fock orbital, ϕ_p , we use the values for the atomic basis functions as a reference point. The ultimate goal is to obtain sets of orbitals where the σ_2^p values are of comparable size with those of the AO basis functions. However, obtaining a tail spread as low for the Hartree–Fock orbitals as for the atomic orbitals is not possible due to the constraints of the Hartree–Fock orbitals, which will be discussed in section 5.

2.5. Reference Point: Projected Atomic Orbitals

When a set of local orbitals that span the virtual Hartree–Fock space is needed, projected atomic orbitals (PAOs) have often been used.^{53–56} The PAOs are generated by projecting the occupied orbital space out of the AO basis

$$|\mu^P\rangle = (1 - \sum_i |i\rangle\langle i|) |\mu\rangle \quad (10)$$

where $|i\rangle$ denote an occupied Hartree–Fock orbital. The occupied Hartree–Fock orbitals are a linear combination of the atomic orbitals

$$|i\rangle = \sum_{\mu} |\mu\rangle C_{\mu i} \quad (11)$$

where $C_{\mu i}$ are the MO coefficients satisfying

$$\langle ij\rangle = \sum_{\mu\nu} C_{\mu i} S_{\mu\nu} C_{\nu j} = \delta_{ij} \quad (12)$$

where

$$S_{\mu\nu} = \langle \mu|\nu \rangle \quad (13)$$

is the AO overlap. The PAOs are thus a redundant set of nonorthogonal orbitals spanning the space of the virtual orbitals. Inserting eq 11 in eq 10, we obtain

$$|\mu^P\rangle = |\mu\rangle - \sum_{\nu\rho} |\nu\rangle D_{\nu\rho} S_{\rho\mu} \quad (14)$$

where the PAO is expressed in terms of the AO overlap matrix and the Hartree–Fock density matrix in the AO basis

$$D_{\mu\nu} = \sum_i C_{\mu i} C_{\nu i} \quad (15)$$

From eq 14, we see that a PAO ($|\mu^P\rangle$) is an AO ($|\mu\rangle$) with a correction that depends on the density matrix elements through the AO overlap matrix. Hence, $|\mu^P\rangle$ will only be local if the

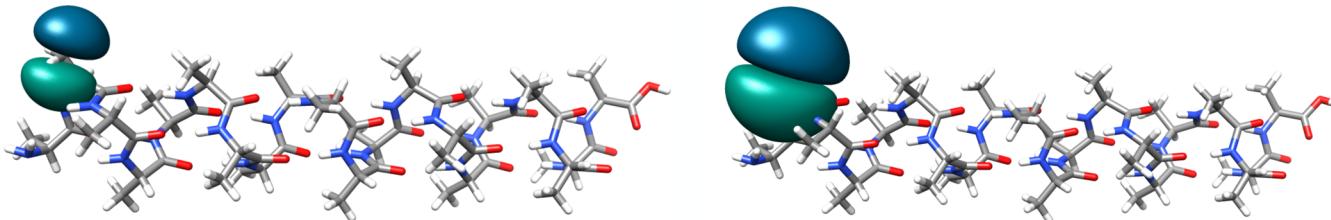


Figure 2. Most diffuse p-function for a carbon atom for the cc-pVTZ basis set ($\sigma_2 = 3.2$ au, $\sigma_4 = 3.5$ au, and $\beta = 1.1$) plotted using contour = 0.03 (left) and contour = 0.003 (right).

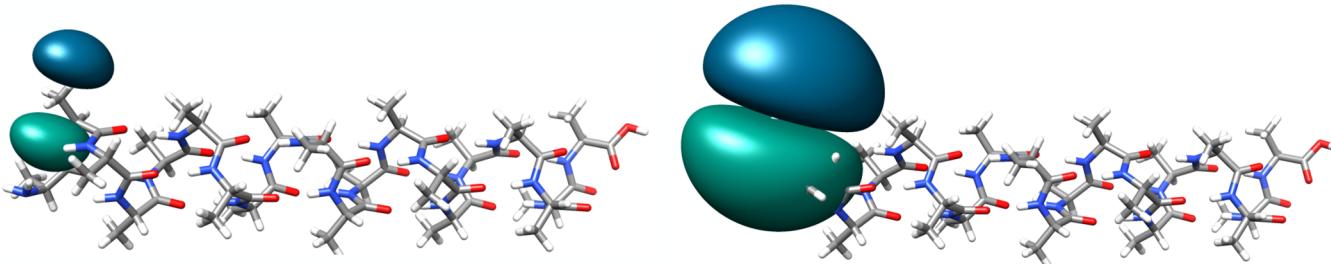


Figure 3. Most diffuse p-function for a carbon atom for the aug-cc-pVTZ basis set ($\sigma_2 = 5.9$ au, $\sigma_4 = 6.4$ au, and $\beta = 1.1$) plotted using contour = 0.03 (left) and contour = 0.003 (right).

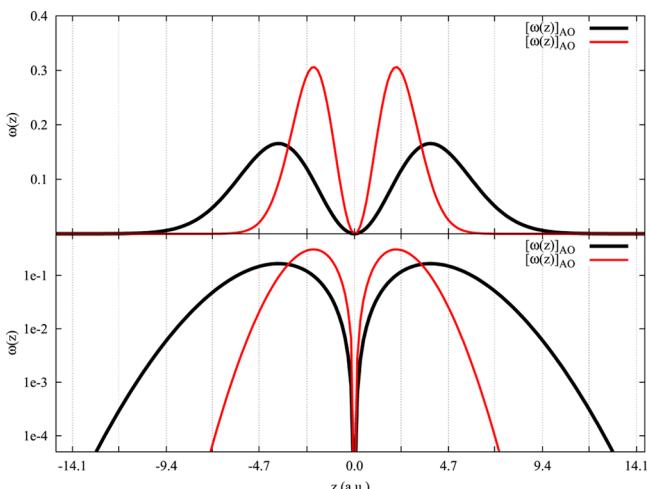


Figure 4. Density distribution function of the most diffuse p-function for a carbon atom for the cc-pVTZ basis set (red) and the most diffuse p-function for a carbon atom in an aug-cc-pVTZ basis set (black) plotted on a regular scale (top) and a log-scale (bottom). The z-axis scale is chosen to make discussion in section 6 more transparent.

Hartree–Fock density matrix is local, whereas the PAOs will have a poor locality for systems with a nonlocal density matrix. Since the PAOs depend on the density matrix and thus on the individual molecular system, the discussion about the locality of the PAOs will be deferred to our discussion about the locality of a set of virtual Hartree–Fock orbitals.

3. ORBITAL LOCALIZATION FUNCTIONS

In this section, we describe the most commonly used localization functions such as the Pipek–Mezey, Boys, and Edmiston–Ruedenberg schemes, together with recently introduced localization functions based on minimizing powers of the orbitals' second and fourth central moments.

3.1. Pipek–Mezey Localization Function

Pipek and Mezey introduced a localization function¹⁵ that measures the number of atomic centers over which a molecular orbital extends.⁹² The Pipek–Mezey localization function is a sum of the squared Mulliken charges for a set of N orthonormal occupied or virtual HF orbitals $\{|p\rangle\}$, and is given by

$$\xi_{PM} = \sum_{p=1}^N \sum_A (Q_{pp}^A)^2 \quad (16)$$

The sum over A runs over atomic centers in the molecule, and Q_{pp}^A is the population of orbital $|p\rangle$ on atomic center A. Q_{pp}^A is defined as

$$Q_{pp}^A = \sum_{\mu \in A} \langle p | \hat{P}_\mu | p \rangle \quad (17)$$

where the projector \hat{P}_μ in Mulliken orbital population frame¹¹⁹ is given as,

$$\hat{P}_\mu^M = \frac{1}{2} (|{}^b\mu\rangle\langle\mu l| + |{}^b\mu\rangle\langle l\mu|) \quad (18)$$

where the functions $\{|{}^b\mu\rangle\}$ are biorthonormal to the atomic basis functions $\{|\mu\rangle\}$, $\langle {}^b\mu | \nu \rangle = \delta_{\mu\nu}$, and are given by

$$|{}^b\mu\rangle = \sum_\nu |\nu\rangle (\mathbf{S}^{-1})_{\nu\mu} \quad (19)$$

where \mathbf{S} is the AO overlap matrix in eq 13. The function, ξ_{PM} , is maximized to obtain a set of local molecular orbitals.

The Mulliken population analysis suffers from some unphysical behavior since the individual Mulliken charges may have occupation numbers greater than 1 or less than 0. This unphysical behavior is due to the fact that in a Mulliken population analysis, overlap populations occur since the AO basis is not orthogonal, and the overlap population is divided equally between the atomic centers, ignoring that different types of atoms have different electronegativity. The unphysical behavior increases when the AO basis set increases in size,¹²⁰

and a more thorough discussion on the unphysical behavior is given by Mulliken.¹²¹

The Pipek–Mezey localization function may also be defined in terms of a Löwdin population,¹²² where the projector \hat{P}_μ in eq 17 is given as

$$\hat{P}_\mu^L = |\bar{\mu}\rangle\langle\bar{\mu}| \quad (20)$$

with

$$|\bar{\mu}\rangle = \sum_\nu |\nu\rangle (\mathbf{S}^{-1/2})_{\nu\mu} \quad (21)$$

When the Löwdin population scheme is used, overlap populations do not occur since the AO basis in eq 21 has been orthogonalized. The population of orbital $|p\rangle$ on center A in eq 17 is therefore always between 0 and 1 for \hat{P}_μ^L of eq 20. For both Löwdin and Mulliken population analysis, the populations of eq 17 are basis set dependent, although less so for the Löwdin populations than the Mulliken populations.¹²³ A discussion of basis set dependence of Mulliken population is given by, for example, Kern et al.¹²⁴ and Politzer et al.¹²⁵ The locality of the Pipek–Mezey occupied and virtual orbitals have been investigated using both the Mulliken and Löwdin populations.³⁵ The occupied orbitals were similar for both population analysis, and the locality of the localized orbitals showed little basis set dependence. The locality of the Pipek–Mezey localized orbitals decrease with an increasing delocalization of the Hartree–Fock density matrix.⁹⁹ For the virtual orbitals, the Löwdin populations based Pipek–Mezey localization was shown to be less basis set dependent than the Mulliken population based counterpart.³⁵ Here, we will therefore only consider the Pipek–Mezey localization implemented using the Löwdin populations. For a thorough discussion on the use of various partial charge estimates in the Pipek–Mezey localization function, we refer to the article by Lehtola et al.⁸⁹ and references therein.

3.2. Edmiston–Ruedenberg Localization Function

Edmiston and Ruedenberg introduced a localization scheme which maximizes the sum of orbital self-repulsion energies

$$\xi_{ER} = \sum_p \langle pp| \frac{1}{r_{12}} |pp\rangle \quad (22)$$

where

$$\langle pp| \frac{1}{r_{12}} |pp\rangle = \int \int \phi_p(\mathbf{r}_1) \phi_p(\mathbf{r}_2) \frac{1}{r_{12}} \phi_p(\mathbf{r}_1) \phi_p(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (23)$$

The ξ_{ER} orbitals have been used in different theoretical investigations,^{41–46} as well as in a more fundamental analysis of the origin of molecular bonding.¹²⁶ The Edmiston–Ruedenberg has a fifth-order computational scaling compared to a third-order scaling for the Boys and Pipek–Mezey schemes, and has, despite a reduction to asymptotically third-order scaling,^{23,27} been used less extensively, and we will not consider the Edmiston–Ruedenberg localization function further.

3.3. Boys Localization Function

The Boys localization^{6–8} is based on minimizing the sum of the orbitals' second central moment

$$\xi_1^{SM} = \sum_p \langle p|(\hat{\mathbf{r}} - \langle p|\hat{\mathbf{r}}|p\rangle)^2|p\rangle = \sum_p \mu_2^p \quad (24)$$

where μ_2^p in eq 1 has been extended to three dimensions. The localization function in eq 24 is known as the Boys–Foster or Boys localization function, but it is a small modification of the original Boys localization function. The modification was proposed by Edmiston and Ruedenberg,⁸ by replacing the $\frac{1}{r_{12}}$ operator in $\langle pp| \frac{1}{r_{12}} |pp\rangle$ of eq 23 by the $(\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2)^2$ operator, followed by a minimization.

$$\begin{aligned} & \sum_p \langle pp|(\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2)^2|pp\rangle \\ &= \sum_p (\langle p|\hat{\mathbf{r}}^2|p\rangle \langle p|p\rangle + \langle p|p\rangle \langle p|\hat{\mathbf{r}}^2|p\rangle - 2\langle p|\hat{\mathbf{r}}|p\rangle \langle p|\hat{\mathbf{r}}|p\rangle) \\ &= 2 \sum_p \langle p|(\hat{\mathbf{r}} - \langle p|\hat{\mathbf{r}}|p\rangle)^2|p\rangle \end{aligned} \quad (25)$$

The set of orbitals obtained from minimizing eq 24 is a set in which the orbitals on average are local. However, we note that a good average locality may be obtained for a set of orbitals even if outliers are present (see section 4.2). The Boys localization function has been widely used in chemistry^{127–134} and is also used in solid state theory for the localization of Wannier functions, for which we refer to the review by Marzari and Vanderbilt¹¹³ and references therein.

3.4. Powers of the Second Central Moment

To produce a set of orbitals where outliers have a less prominent role than in a Boys localization, Jansík et al.³² imposed a penalty for orbitals with a large second central moment by minimizing powers of the second central moment rather than just the second central moment itself

$$\xi_m^{SM} = \sum_p (\mu_2^p)^m \quad (26)$$

where μ_2^p is defined by eq 24 and m is a positive integer. Increasing the power m increases the penalty on orbitals with large μ_2^p values, since powers of μ_2^p affect the function value more severely than μ_2^p itself. Recall that $m = 1$ yields the Boys localization function. Jansík et al. have shown³² that it is important to use a power $m = 2$ but that increasing the power beyond $m = 2$ does not have any significant effect on the locality of the orbital set. We will for that reason only report ξ_m^{SM} localization results. Recently, Liu et al.³⁸ have used powers less than one for the Boys localization function and concluded that powers less than one give better locality than powers greater than one. However, they did not separate the localization of the core orbitals from the localization of valence orbitals in the least-change procedure and thereby introduced an extensive mixing of the core and the valence orbitals. For this reason, the obtained orbitals cannot be used in frozen core calculations and in local correlation methods,¹¹² and we will not consider these orbitals.

3.5. Powers of the Fourth Central Moment

The second central moment weight function targets the bulk of the orbital (see section 2). However, for local orbitals, it is also important to have a rapid tail decay. Following the argumentation for using the fourth moment weight function to increase the weighting of the tails, see section 2.2, the fourth central moment may be used as a localization function to obtain orbitals with rapidly decaying tails.³⁴ Extending eq 5 to three dimensions, we write the fourth central moment as,

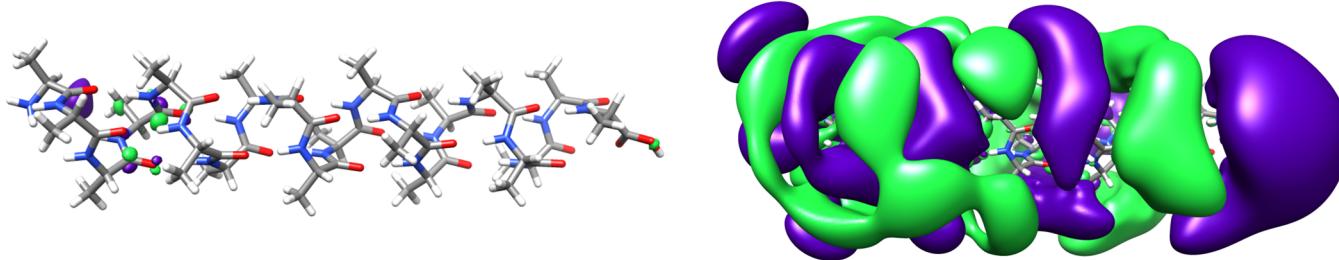


Figure 5. A canonical MO with $\sigma_2 = 24.8$ au for alanine(20) in the aug-cc-pVTZ basis, plotted for contour = 0.03 (left) and contour = 0.003 (right), as an illustration of a nonlocal orbital.

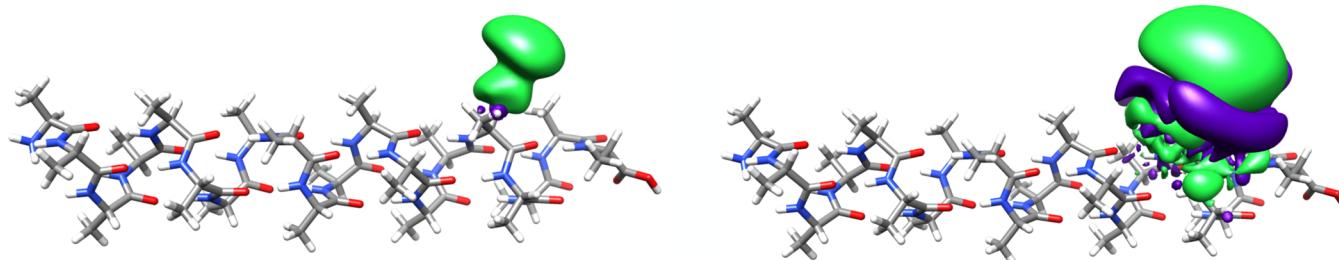


Figure 6. ξ_2^{FM} orbital with $\sigma_2 = 4.1$ au, and $\sigma_4 = 5.3$ a.u ($\beta = 1.3$) for alanine(20) in the aug-cc-pVTZ basis plotted on contour = 0.03 (left) and contour = 0.003 (right).

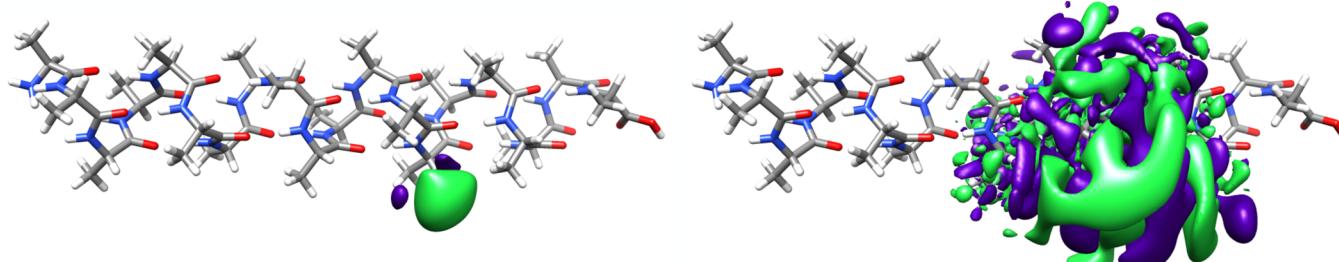


Figure 7. ξ_1^{SM} orbital with $\sigma_2 = 4.1$ au and $\sigma_4 = 7.0$ a.u ($\beta = 1.6$) for alanine(20) in an aug-cc-pVTZ basis plotted on contour = 0.03 (left) and contour = 0.003 (right).

$$\mu_4^p = \langle p | (\hat{r} - \langle p | \hat{r} | p \rangle)^4 | p \rangle \quad (27)$$

As was the case for the second central moment, outliers may be removed using powers in the localization function,

$$\xi_m^{\text{FM}} = \sum_p (\mu_4^p)^m \quad (28)$$

where the positive integer m imposes a penalty for outliers. As for powers of the second central moment localization function, a power two is sufficient to remove outliers for the fourth moment localization function,³⁴ and results will only be reported for the ξ_2^{FM} localization function.

4. EXEMPLIFYING ORBITAL LOCALITY MEASURES

In this subsection, we give some specific examples of Hartree–Fock orbitals for alanine(20) (an α -helix consisting of 20 alanine residues, $C_{60}N_{20}O_{21}H_{102}$) in the aug-cc-pVTZ basis, to illustrate the locality measures introduced in section 2. The discussion is separated into describing the locality of an individual orbital and describing the locality of a set of orbitals.

4.1. Locality of Individual Orbitals

The locality of an individual orbital may be understood in terms of a pictorial contour representation or in terms of the locality measures σ_2 , σ_4 , and β . To illustrate the relationship between

the pictorial and the numerical locality measures, we consider a canonical molecular orbital (CMO), a Boys localized orbital (ξ_1^{SM}), and a fourth moment localized orbital (ξ_2^{FM}). The ξ_1^{SM} and ξ_2^{FM} localization functions were described in section 3.

The CMO plotted in Figure 5 has $\sigma_2 = 24.8$ au, $\sigma_4 = 26.8$ au, and $\beta = 1.1$. Both the σ_2 value and the plot emphasize that the orbital is nonlocal. The σ_4 and β values are of little interest for nonlocal orbitals, where the orbital stretches across the entire molecular system.

Next, we consider a ξ_2^{FM} virtual orbital ($\phi_{\text{FM}2}$) and a ξ_1^{SM} virtual orbital ($\phi_{\text{SM}1}$) where both orbitals have the same second moment spreads, $\sigma_2 = \sigma_2^{\text{FM}2} = \sigma_2^{\text{SM}1} = 4.1$ au, but the orbitals' fourth moment spreads differ ($\sigma_4^{\text{FM}2} = 5.3$ au, $\beta_{\text{FM}2} = 1.3$ versus $\sigma_4^{\text{SM}1} = 7.0$ au, $\beta_{\text{SM}1} = 1.6$). $\phi_{\text{FM}2}$ and $\phi_{\text{SM}1}$ are plotted using contour = 0.03 and contour = 0.003, in Figures 6 and 7, respectively. From Figures 6 and 7, left (contour = 0.03), there is no apparent difference between the two orbitals in terms of spatial extent. In contrast, if we look at Figures 6 and 7, right (contour = 0.003), the orbital in Figure 7 clearly has a larger spatial extent than the orbital in Figure 6. σ_2 is not able to detect this difference in spatial extent, but the difference is reflected by the different σ_4 values. For orbitals with the same σ_2 comparing their σ_4 values will thus give information about the tail decay of the orbitals relative to each other. For orbitals

where there is some difference in the σ_2 values, β serves as a normalized measure, making it possible to compare tail decay of orbitals with different σ_2 values.

4.2. Locality of a Set of Orbitals

When localizing Hartree–Fock orbitals, we are concerned with the locality of all the orbitals in the set. Orbital set locality may be characterized in different ways. We may report the average locality of the orbital set in terms of

$$\sigma_2^{\text{avg}} = \frac{1}{N} \sum_{p=1}^N \sigma_2^p \quad (29)$$

$$\sigma_4^{\text{avg}} = \frac{1}{N} \sum_{p=1}^N \sigma_4^p \quad (30)$$

where N is the number of orbitals in the set. The measures in eqs 29 and 30 give an indication on whether the set as a whole is local or not. Large σ_2^{avg} and σ_4^{avg} values thus mean that the set is not local. Contrary, if the σ_2^{avg} and σ_4^{avg} values are small, the set as a whole is not necessarily local. A good average locality may hide that a single or several orbitals are considerably less local than the set on average. Orbitals, $\phi_p(r)$, for which $\sigma_2^p \gg \sigma_2^{\text{avg}}$ and $\sigma_4^p \gg \sigma_4^{\text{avg}}$ are termed outliers. A better characterization of the locality of the orbital set is therefore to report the locality of the least local orbital in the set (“the weakest link”). This may be done by using maximum values (i.e., report the largest σ_2 , σ_4 , and β found in the orbital set of N orbitals)

$$\sigma_2^{\text{max}} = \max(\sigma_2^1, \sigma_2^2, \dots, \sigma_2^N) \quad (31)$$

$$\sigma_4^{\text{max}} = \max(\sigma_4^1, \sigma_4^2, \dots, \sigma_4^N) \quad (32)$$

$$\beta^{\text{max}} = \max(\beta^1, \beta^2, \dots, \beta^N) \quad (33)$$

To demonstrate that σ_2^{max} and σ_4^{max} may be used as good measures of set locality, consider the ξ_1^{SM} (Boys) localized virtual orbitals for alanine(20) in the aug-cc-pVTZ basis. The ξ_1^{SM} orbitals have $\sigma_2^{\text{avg}} = 2.2$ au and $\sigma_4^{\text{avg}} = 3.7$ au. The σ_2^{avg} and σ_4^{avg} values are small, indicating that the set is local. However, plotting the σ_2 and σ_4 values for all the 5689 virtual orbitals in the set (Figure 8), it is seen that many of the orbitals have large

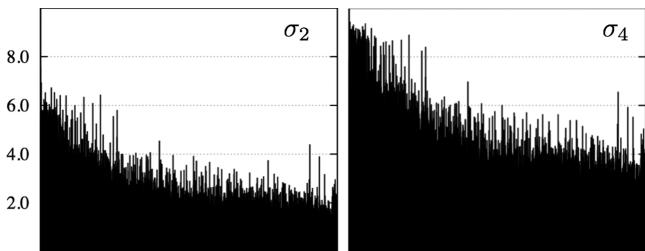


Figure 8. Second and fourth moment orbital spreads (in au) for all 5689 virtual orbitals of alanine(20) in the aug-cc-pVTZ basis localized using the ξ_1^{SM} localization function. $\sigma_2^{\text{avg}} = 2.2$ au, $\sigma_4^{\text{avg}} = 3.7$ au, $\sigma_2^{\text{max}} = 7.6$ au, and $\sigma_4^{\text{max}} = 9.4$ au.

σ_2 and σ_4 values. The maximum values $\sigma_2^{\text{max}} = 7.6$ au and $\sigma_4^{\text{max}} = 9.4$ au are thus large compared to the average values, and from Figure 8 we see that the orbital set contains many outliers.

If average and maximum orbital spread values are similar, the orbitals in the set have approximately the same locality. To illustrate this, we consider the σ_2 and σ_4 values for the virtual orbitals of alanine(20) localized using ξ_2^{FM} . The average values

are $\sigma_2^{\text{avg}} = 2.7$ au and $\sigma_4^{\text{avg}} = 4.5$ au, while the maximum values are $\sigma_2^{\text{max}} = 4.1$ au and $\sigma_4^{\text{max}} = 5.3$ au. The locality of all orbitals are plotted in Figure 9. For orbital sets where there is little

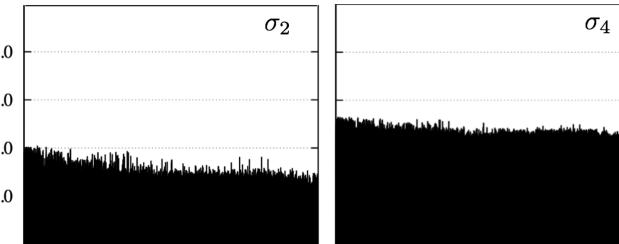


Figure 9. Second and fourth moment orbital spreads (in au) for all 5689 virtual orbitals of alanine(20) in a aug-cc-pVTZ basis localized using the ξ_2^{FM} localization function. $\sigma_2^{\text{avg}} = 2.7$ au, $\sigma_4^{\text{avg}} = 4.5$ au, $\sigma_2^{\text{max}} = 4.1$ au, and $\sigma_4^{\text{max}} = 5.3$ au.

difference between the average and maximum orbital spread values, the plot containing all orbital spread values has a definite baseline, as seen in Figure 9. Contrary, if there is a large difference between the average and maximum orbital spread values, there is no definite baseline of locality, as seen in Figure 8.

5. CONSTRAINTS ON THE LOCALIZABILITY OF A SET OF MOLECULAR ORBITALS

The Hartree–Fock optimization condition defines a physical separation of the orbital space into an occupied and a virtual part. For the occupied orbital space, the localized orbitals further have to generate the orbital invariant Hartree–Fock density matrix in eq 15. The localizability of a set of occupied orbitals is therefore connected to the locality of the Hartree–Fock density matrix and thus to the electronic structure of the molecular system. We denote the locality limitation that is introduced by the Hartree–Fock optimization condition a physical limitation on the localizability. The localizability of the Hartree–Fock orbitals is also limited by the requirement that the orbitals are orthogonal. This is a mathematical constraint, which is imposed on the orbitals that ultimately leads to orbital tails because the orbitals have to have a complicated nodal structure to fulfill the orthogonality conditions. In this section, we discuss the implications these restrictions have on the localizability of the orbitals, following a discussion by Høyvik et al.⁹⁹ In particular, we review the discussion whether it is the physical or the mathematical restriction that is the main source of the tails. Further, we look at how the locality of the electronic structure of a molecule, as measured by the locality of the Hartree–Fock density matrix, affects the localizability.

5.1. Orbital Bases for Separating the Physical and Mathematical Constraints on Locality

We now discuss how the localizability of a set of Hartree–Fock orbitals is affected by the mathematical and the physical constraints. To separate the orbital orthogonality constraint from the Hartree–Fock optimization condition, we will consider the localizability of three orbital bases that all span the same orbital space but exhibit none, one, or both of these constraints.⁹⁹ These bases are (i) AOs $\{\langle \mu | \rangle\}$ which constitute a basis that is nonorthogonal and does not satisfy an occupied-virtual partitioning of the orbital space, (ii) AOs that are orthogonalized and localized (LOAOs) $\{\langle \mu^L | \rangle\}$ and thus constitute a basis which is orthogonal and does not satisfy an

occupied-virtual partitioning of the orbital space, and (iii) localized Hartree–Fock orbitals (LMOs, $\{|p\rangle\} = \{|i\rangle, |a\rangle\}$) that constitutes a basis which is orthogonal and satisfies an occupied-virtual partitioning of the orbital space. We use indices i,j,\dots for occupied orbitals and indices a,b,\dots for virtual orbitals, whereas P,Q,\dots are general indices which may denote Hartree–Fock orbitals, AOs, or LOAOs. Hence, we have three orbital bases ($\{|\mu\rangle\}$, $\{|\bar{\mu}^L\rangle\}$, and $\{|p\rangle\}$) which represent three different scenarios with respect to the optimization condition and the orthogonality constraint. The locality of these orbital bases will be expressed in terms of σ_2^{\max} , σ_4^{\max} , and β^{\max} of eqs 31, 32, and 33.

To obtain more details on the locality of the orbitals, the orbital spread information will be supplemented by information that describes how the density distributions, $\phi_P(r)$ $\phi_Q(r)$, between orbitals $\phi_P(r)$ and $\phi_Q(r)$ decay as a function of the distance R_{PQ} where

$$R_{PQ} = |\langle \phi_p | \hat{r} | \phi_p \rangle - \langle \phi_Q | \hat{r} | \phi_Q \rangle| \quad (34)$$

is the distance between the orbital centers. Since the overlap between two orthonormal orbitals by construction is 1 or 0, the overlap cannot be used to characterize the density distributions. However, the numerical overlap Ω_{PQ} may be used as a measure of the magnitude of the density distribution

$$\Omega_{PQ} = \int |\phi_p(r)| |\phi_Q(r)| dr \quad (35)$$

The numerical overlap Ω_{PQ} will be plotted as a function of R_{PQ} in terms of the largest Ω_{PQ} for a given distance. For simplicity, we will only report plots that contain the maximum Ω_{PQ} for a given distance [denoted $\max(\Omega_{PQ})$ plots].

5.2. Numerical Illustrations for Separating the Physical and Mathematical Effects

Following Høyvik et al.,⁹⁹ we will now analyze the locality of the Hartree–Fock orbitals for a diamond structure ($C_{331}H_{216}$) and a graphene sheet ($C_{106}H_{28}$). Diamond is an example of a molecule that is an insulator and has a localized electronic structure, while graphene represents an example of a delocalized system. For diamond and graphene, the effect of the physical and mathematical constraint on the localizability is examined by reporting the locality of the three bases (AOs, LOAOs, and LMOs) in terms of maximum orbital spreads. The calculations employ the cc-pVDZ basis for both diamond and graphene.

We first consider the different locality of the electronic structure of diamond and graphene as expressed in terms of the locality of the density matrix in the AO basis (eq 15). To visualize the locality, we have in Figure 10 plotted the distance decay of the density matrix where the values reported in Figure 10 are the max $|D_{\mu\nu}|$ as a function of the distance, $R_{\mu\nu}$, with spacings of 1.0 au, between the centers of the atomic orbitals $|\mu\rangle$ and $|\nu\rangle$ (eq 34). The decay of the density matrix reflects the local character of the electronic structure for diamond with a monotonic decrease of the density matrix elements as the distance $R_{\mu\nu}$ increase. For graphene, the decrease of the density matrix elements for increasing $R_{\mu\nu}$ is slow beyond 5 Å, in accordance with the nonlocal character of the electronic structure of graphene.

We will now compare the locality of the three different bases (AOs, LOAOs, and LMOs) for diamond and graphene where the LOAOs and LMOs have been localized by minimizing the fourth central moment, ξ_2^{FM} , given in section 3 (eq 28). The

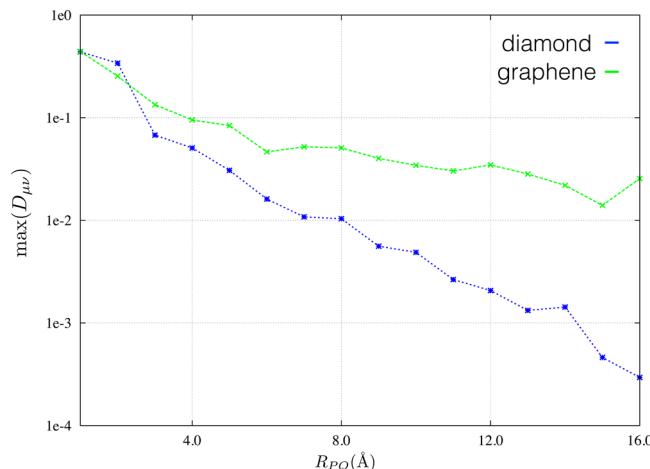


Figure 10. Distance decay of density matrix elements in the AO basis for diamond and graphene.

reason for using this localization function is its robustness, which will be discussed in Sections 6 and 7. In Table 2, we report σ_2^{\max} , σ_4^{\max} , and β^{\max} for the AOs, LOAOs, and LMOs.

Table 2. σ_2^{\max} (au), σ_4^{\max} (au), and β^{\max} Values for the AOs, LOAOs, LMOs, where the LOAOs and LMOs Have Been Localized Using ξ_2^{FM}

	diamond			graphene		
	σ_2^{\max}	σ_4^{\max}	β^{\max}	σ_2^{\max}	σ_4^{\max}	β^{\max}
AOs	2.9	3.1	1.07	2.9	3.1	1.07
LOAOs	2.4	3.4	2.06	2.3	3.3	2.07
LMOs, occupied	1.8	2.5	1.49	2.8	4.1	1.92
LMOs, virtual	2.8	4.0	1.90	3.1	3.9	2.02

Considering the bulk behavior in the different orbital bases, we see from Table 2 that for both diamond and graphene, the σ_2^{\max} values are of similar size for the AOs and virtual LMOs and with slightly smaller values for the LOAOs. For diamond, the σ_2^{\max} value is reduced by about one-third for the LMOs occupied compared to the AOs, while for graphene σ_2^{\max} is of similar size for the occupied LMOs and the AOs. The large decrease of the σ_2^{\max} value for diamond may be attributed to the local character of the electronic structure of diamond.

Considering the tail behavior of the different orbital bases, we see from Table 2 that the AOs have β^{\max} value close to one, while the β^{\max} values for all other orbital bases are larger than one. All orbital bases except the AOs therefore have a significant density distribution in the tail region. From a conceptual point-of-view, this is easy to understand, since the LOAOs and the LMOs are orbitals expanded in the basis of AOs, and therefore, the LOAOs and LMOs will contain contributions from AOs on different atomic centers. Contrary, the AOs are gaussians centered on single atoms and thus have no tail since the gaussians have no tails (see Table 1).

To get more insight into the bulk and tail behavior of the different orbital bases, we report in Figure 11 for diamond (left) and graphene (right) $\max(\Omega_{PQ})$ plots for the AOs, LOAOs, and LMOs. Looking at the numerical overlaps in Figure 11, we see in accordance with the σ_2^{\max} values of Table 2 that for small R_{PQ} , the numerical overlap values for the LOAOs are slightly below the numerical overlap values for the AOs. Further, for diamond, the occupied LMOs for small R_{PQ} have numerical overlap

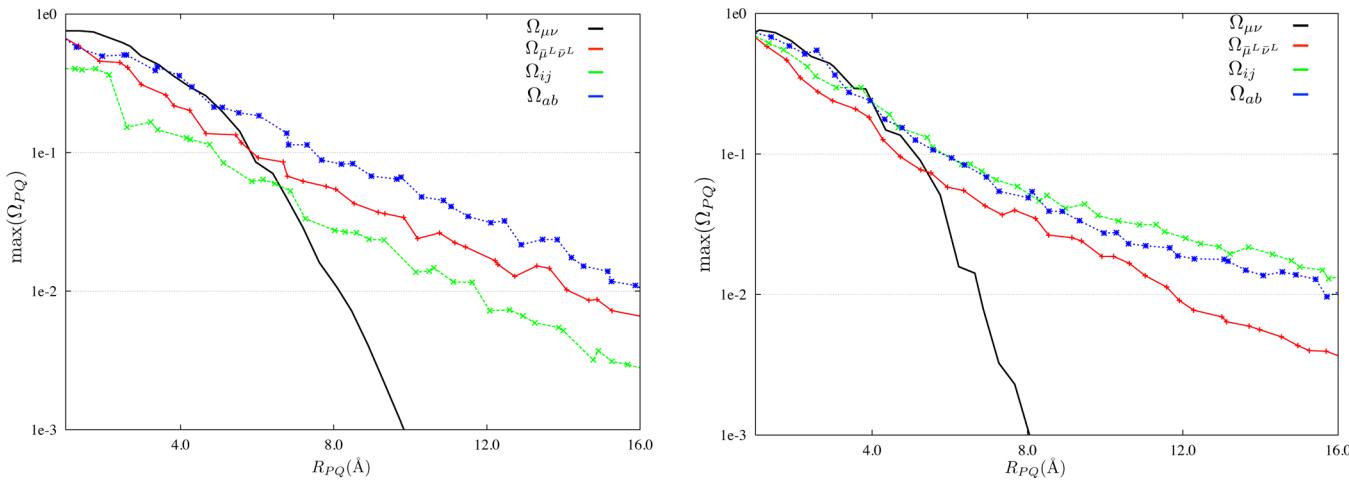


Figure 11. Numerical overlaps $\Omega_{\mu\nu}$ (black), $\Omega_{\bar{\mu}^L \bar{\nu}^L}$ (red), Ω_{ij} (green), and Ω_{ab} (blue) plotted as a function of the distance between orbital centers for diamond (left) and graphene (right).

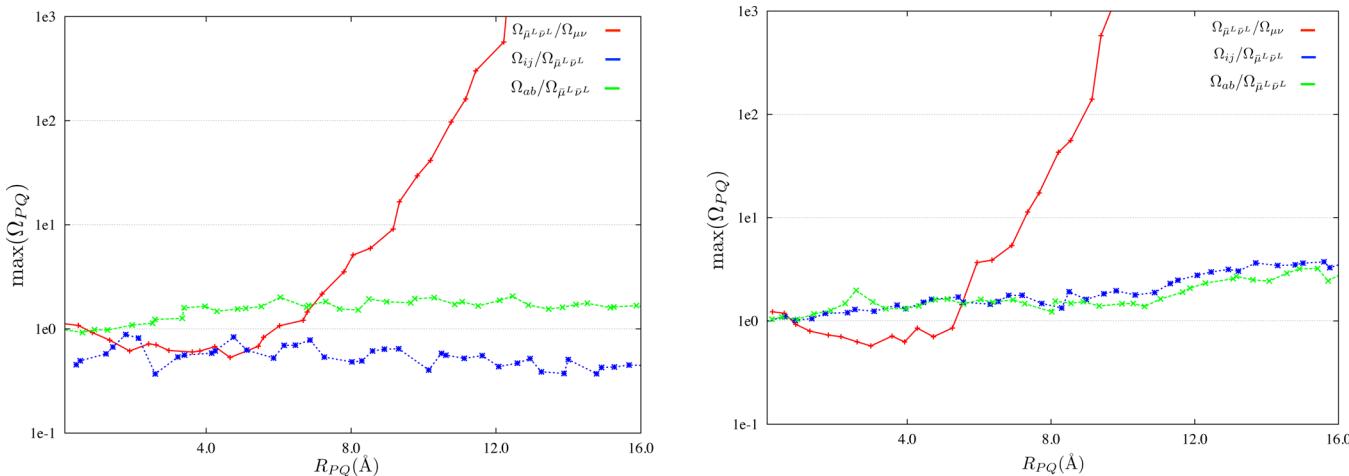


Figure 12. Ratios $\frac{\Omega_{\bar{\mu}^L \bar{\nu}^L}}{\Omega_{\mu\nu}}$ (red), $\frac{\Omega_{ij}}{\Omega_{\bar{\mu}^L \bar{\nu}^L}}$ (green), and $\frac{\Omega_{ab}}{\Omega_{\bar{\mu}^L \bar{\nu}^L}}$ (blue) plotted as a function of distance for diamond (left) and graphene (right).

values that are below the values for the LOAOs. For the occupied LMOs of graphene and the virtual LMOs for both diamond and graphene, the numerical overlap values are about superimposed on the ones for the LOAOs. Thus, it is possible to obtain a locality of the molecular orbitals of diamond and graphene which is similar to the LOAOs, despite the occupied-virtual partitioning of the space. Considering the tail region, the numerical overlap values in Figure 11 confirm that the AOs have no tail while for all the other bases we have a significant density distribution in the tail region.

We will now consider the effects of the orthogonalization and of the occupied-virtual orbital space splitting separately, plotting in Figure 12 the distance dependence of the size of $\Omega_{\bar{\mu}^L \bar{\nu}^L}$ relative to $\Omega_{\mu\nu}$ (red curve), the size of Ω_{ij} relative to $\Omega_{\bar{\mu}^L \bar{\nu}^L}$ (green curve), and the size of Ω_{ab} relative to $\Omega_{\bar{\mu}^L \bar{\nu}^L}$ (blue curve). The red curve explicitly describes the effect of the orthogonalization on the distance decay of density distributions, while the green and blue curves describes the effect of requiring the optimization condition to be fulfilled. From Figure 12, it is seen that the red curve is dominating beyond 6–7 Å for both diamond (left) and graphene (right). Hence, already for rather small separations the orthogonalization effect is seen to be strong and the distance decay of orbital density distributions for larger distances are thus governed by the

orthogonalization requirement and not by the physical separation of the orbital space into an occupied and a virtual part.

6. LOCALITY ILLUSTRATIONS FOR SINGLE ORBITALS

In this section, we examine characteristics of individual orbitals that are obtained by localizing sets of orbitals using the localization functions of section 3. For simplicity, we will consider arachidic acid ($C_{20}O_2H_{40}$), a one-dimensional molecular system, displayed in Figure 13, with the molecular axis in the z -direction and using the aug-cc-pVTZ basis.



Figure 13. Arachidic acid molecule.

The locality of an individual orbital will be expressed in terms of second and fourth moment orbital spreads and orbital density distributions along the z -axis (eq 9). For each localization function, we will plot $\omega_p(z)$ for both an occupied and a virtual orbital, where these orbitals are chosen among the least local orbitals in the set. Further, for the virtual orbitals, we

will also plot the corresponding normal distribution, $N(z, [\sigma_2^p]_z)$, since the structure of $\omega(z)$ for the virtual orbitals is rather complex. The normal distributions corresponding to the $\omega_p(z)$ distributions are obtained from eq 4 using the z -component $[\sigma_2^p]_z$ of σ_2^p

$$[\sigma_2^p]_z = \sqrt{\langle p|\hat{z}^2|p\rangle - \langle p|\hat{z}|p\rangle^2} \quad (36)$$

6.1. Occupied Orbitals

The orbital density distributions for occupied orbitals (among the least local) of ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} localized orbitals are denoted $[\omega(z)]_{\text{PM}}$, $[\omega(z)]_{\text{SM1}}$, $[\omega(z)]_{\text{SM2}}$, and $[\omega(z)]_{\text{FM2}}$, respectively. The distributions are plotted in Figure 14 on a

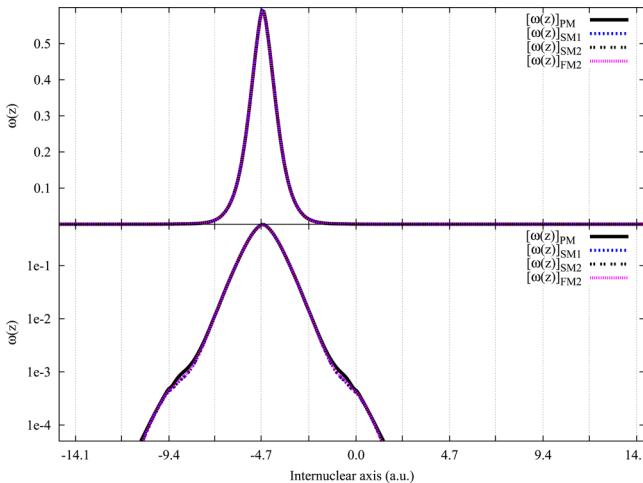


Figure 14. $[\omega(z)]_{\text{PM}}$, $[\omega(z)]_{\text{SM1}}$, $[\omega(z)]_{\text{SM2}}$, and $[\omega(z)]_{\text{FM2}}$ for occupied orbitals for arachidic acid in the aug-cc-pVTZ basis, plotted on a regular scale (top) and a log-scale (bottom). The vertical gridlines indicate positions of the carbon atoms. The z -axis range is chosen to be consistent with the range used in Figure 4.

regular scale (top) and a log-scale (bottom). Figure 14 shows that $\omega(z)$ for the different localization schemes are indistinguishable from each other, with a simple shape of the orbital density distributions and a rapid distance decay. The reason for this similarity is the simple bonding structure of arachidic acid (see Figure 13). The scale on the z -axis in Figure 14 is the same as the scale on the axis in Figure 4. Figure 14 shows that for $\omega(z) > 10^{-4}$ (see Figure 14, bottom) the distribution only encompasses five carbon centers (marked by the vertical gridlines), while about twice this number is encompassed by the most diffuse p-function for a carbon in aug-cc-pVTZ (see Figure 4). However, the rapid decay behavior found for the occupied orbitals in Figure 14 does not occur for occupied orbitals of molecules with a more complex bonding structure for which the occupied orbitals display a less localized character (e.g., the graphene sheet, see Figure 10 and the discussion in section 5.2).

6.2. Virtual Orbitals

The density distributions $[\omega(z)]_{\text{PM}}$, $[\omega(z)]_{\text{SM1}}$, $[\omega(z)]_{\text{SM2}}$, and $[\omega(z)]_{\text{FM2}}$ (following the notation of section 6.1) for the least local virtual orbitals are plotted in Figure 15 (top) on a regular scale and Figure 16 (top) on a log-scale. For comparison, $\omega(z)$ has also been plotted for a projected atomic orbital (PAO), $[\omega(z)]_{\text{PAO}}$. The normal distributions (eq 4) for the orbitals, generated from the z -components of the second moment

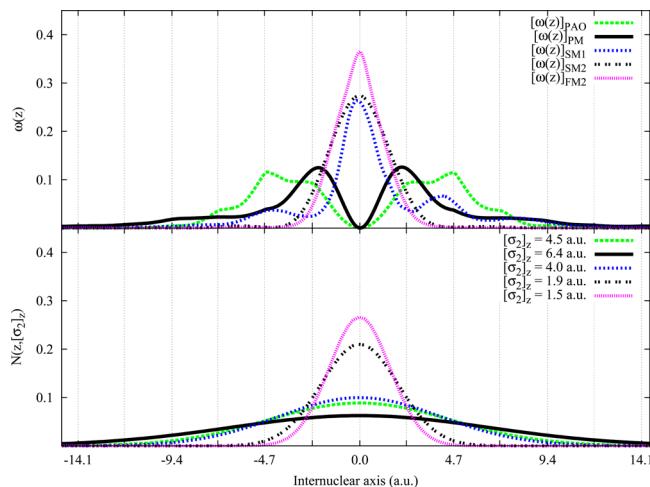


Figure 15. Top: $[\omega(z)]_{\text{PM}}$, $[\omega(z)]_{\text{SM1}}$, $[\omega(z)]_{\text{SM2}}$, $[\omega(z)]_{\text{FM2}}$, and $[\omega(z)]_{\text{PAO}}$ for the virtual orbitals of arachidic acid in the aug-cc-pVTZ basis. Bottom: normal distributions generated using $[\sigma_2]_z$ of the orbitals plotted in the upper plot. The gridlines indicate nuclear positions for the carbon atoms. The z -axis range is chosen to be consistent with the range used in Figure 4.

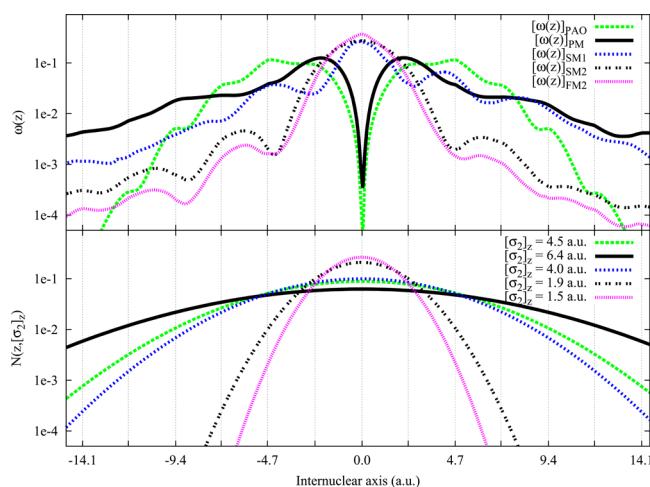


Figure 16. Distributions (top) and normal distributions (bottom) from Figure 15 plotted on a log-scale.

spreads, are plotted in Figures 15 and 16 (bottom). In Table 3, we have listed the second and fourth moment orbital spreads for the plotted orbitals, together with the z -component of their second moment orbital spreads, $[\sigma_2]_z$.

From Figure 15, we see that the structure of $\omega(z)$ depends strongly on which localization function is used to generate the orbital. $[\omega(z)]_{\text{PM}}$ and $[\omega(z)]_{\text{PAO}}$ are split into two peaks of approximately equal height, while $[\omega(z)]_{\text{SM1}}$, $[\omega(z)]_{\text{SM2}}$, and $[\omega(z)]_{\text{FM2}}$ display a single peak. $[\omega(z)]_{\text{PAO}}$ describes a p atomic

Table 3. σ_2 , σ_4 , and $[\sigma_2]_z$ (in au) for the Virtual Orbitals and the PAO Plotted in Figures 15 and 16

scheme	σ_2	σ_4	$[\sigma_2]_z$
PAO	6.4	6.9	4.5
ξ_{PM}	10.0	12.0	6.4
ξ_1^{SM}	6.8	8.7	4.0
ξ_2^{SM}	4.2	6.1	1.9
ξ_2^{FM}	3.6	5.1	1.5

orbital where the occupied orbital space component is projected out. The splitting of $[\omega(z)]_{\text{PAO}}$ into two peaks reflects that the p-character of the atomic orbital is preserved after the occupied space is projected out. The Pipek-Mezey localization function preserves the $\sigma-\pi$ separation of the orbitals,^{15,135} and the least local virtual orbital therefore also has preserved its p-character during the Pipek-Mezey localization and the orbital therefore is split into two peaks. For $[\omega(z)]_{\text{SM1}}$, $[\omega(z)]_{\text{SM2}}$, and $[\omega(z)]_{\text{FM2}}$, there is an extensive mixing between the orbitals, resulting in orbital distribution functions that are more local and dominated by a single peak.

From Figure 15, we also see that even for a simple linear system like arachidic acid, there are large differences in the bulk and tail locality of the virtual orbitals. As expected from the structure of the localization functions, the most acute peak is obtained for $[\omega(z)]_{\text{FM2}}$. The width of $[\omega(z)]_{\text{SM2}}$ is similar to that of $[\omega(z)]_{\text{FM2}}$ in Figure 15; however, the log-scale plot in Figure 16 reveals a faster tail decay for $[\omega(z)]_{\text{FM2}}$. The $[\omega(z)]_{\text{SM1}}$ orbital is dominated by a single peak that has large shoulders (see Figure 15, top), resulting in a quite broad distribution compared to distributions for $[\omega(z)]_{\text{SM2}}$ and $[\omega(z)]_{\text{FM2}}$. A large shoulder as seen for $[\omega(z)]_{\text{SM1}}$ does not occur for $[\omega(z)]_{\text{SM2}}$ and $[\omega(z)]_{\text{FM2}}$ since the localization functions ξ_2^{SM} and ξ_2^{FM} contains a power 2 on the central moments and therefore introduce a penalty for having a distribution with a density located away from the orbital center.

With consideration of tail decays, the localized virtual orbitals $[\omega(z)]_{\text{SM1}}$, $[\omega(z)]_{\text{SM2}}$, and $[\omega(z)]_{\text{FM2}}$ have tails to an extent that depend on the weights of the density distributions in the localization functions used to generate the orbitals. $[\omega(z)]_{\text{PM}}$ and $[\omega(z)]_{\text{PAO}}$ show slow tail decays which are seen also in Figure 15, (top). The locality of the virtual orbitals thus depends strongly on which localization function is used.

7. LOCALITY OF A SET OF ORBITALS

In this section, we consider the locality of sets of occupied and virtual orbitals localized using the localization functions in section 3. The locality measures introduced in section 2 are used to quantify the locality. We report calculations on arachidic acid (aa, see section 6), the I27_{SS} domain ($C_{124}N_7O_{37}S_3H_{192}$) of the titin protein (referred to as titin, for simplicity), alanine(20) (see section 4), insulin ($C_{257}N_{65}O_{77}H_{382}$), and a graphene sheet ($C_{106}H_{28}$). The basis sets used are Dunning's correlation consistent basis sets cc-pVXZ for cardinal numbers X = D,T,Q and Dunning's augmented correlation consistent basis sets aug-cc-pVXZ, X = D,T. For calculations using aug-cc-pVXZ, the cc-pVXZ basis was used for the hydrogen atoms to avoid linear dependencies. In section 7.1, we will discuss the locality of the set of occupied orbitals, while the locality of the sets of virtual orbitals are discussed in section 7.2 for nonaugmented basis sets and in section 7.3 for augmented basis sets.

7.1. Locality of the Occupied Orbitals

For the molecules and basis sets described above, we report in Table 4 the locality of the ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} localized occupied orbitals in terms of σ_2^{max} and σ_4^{max} values.

With the exception of graphene, which has a delocalized electronic structure (see Figure 10), the locality of the occupied orbitals produced by the different localization schemes are similar. The locality is the same whether augmented or nonaugmented basis sets are used. In particular, the maximum orbital spread values are nearly identical for orbitals produced

Table 4. σ_2^{max} and σ_4^{max} Values (in au) for Sets of Occupied Orbitals for Different Combinations of Molecules and Basis Sets

	ξ_{PM}		ξ_1^{SM}		ξ_2^{SM}		ξ_2^{FM}	
	σ_2^{max}	σ_4^{max}	σ_2^{max}	σ_4^{max}	σ_2^{max}	σ_4^{max}	σ_2^{max}	σ_4^{max}
aa/cc-pVDZ	1.8	2.2	1.6	2.2	1.6	2.2	1.6	2.2
aa/cc-pVTZ	1.8	2.2	1.6	2.2	1.6	2.2	1.6	2.2
aa/cc-pVQZ	1.8	2.2	1.6	2.2	1.6	2.2	1.6	2.2
aa/aug-cc-pVDZ	1.8	2.2	1.6	2.2	1.6	2.2	1.6	2.2
aa/aug-cc-pVTZ	1.8	2.2	1.6	2.2	1.6	2.2	1.6	2.2
titin/cc-pVDZ	2.8	3.4	2.2	3.0	2.2	2.9	2.2	2.9
titin/cc-pVTZ	2.9	3.4	2.3	3.0	2.2	3.0	2.2	2.9
titin/aug-cc-pVDZ	2.9	3.5	2.3	3.1	2.2	3.0	2.2	2.9
ala(20)/cc-pVTZ	1.9	2.4	1.7	2.3	1.6	2.2	1.6	2.2
ala(20)/aug-cc-pVTZ	1.9	2.4	1.7	2.3	1.6	2.2	1.6	2.2
insulin/cc-pVDZ	2.8	3.3	2.2	2.9	2.1	2.8	2.2	2.8
graphene/cc-pVDZ	3.8	7.3	3.4	5.6	2.5	5.0	2.6	4.2

by the ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} localization functions, while the values are slightly larger for the ξ_{PM} localized occupied orbitals. The σ_2^{max} values in Table 4 are smaller than the σ_2^{max} for the atomic basis sets in Table 1. All localization functions are thus capable of producing a set of local occupied orbitals with a more confined bulk than the AOs. However, the σ_4^{max} show that the localized occupied orbitals have larger tails than the AOs. For graphene, the σ_2^{max} values ranges from 3.8 au (for ξ_{PM}) to 2.5 au (for ξ_2^{SM}), while the σ_4^{max} ranges from 7.3 au to 4.2 au for ξ_{PM} and ξ_2^{FM} , respectively. A pictorial representation of the large difference between the maximum orbital spread values for a ξ_{PM} and a ξ_2^{FM} localization is given in Figure 17, where the least local orbitals for ξ_{PM} and ξ_2^{FM} (as measured by σ_4) have been plotted. Figure 17 shows that the orbital density for the ξ_{PM} orbital stretches across many more atomic centers than does the ξ_2^{FM} orbital density. ξ_{PM} is thus not capable of producing a set of local occupied orbitals for a system with a delocalized electronic structure. Contrary, σ_2^{max} of 2.6 au for the ξ_2^{FM} localized orbitals for graphene is in the same range as σ_2^{max} values for the systems with a localized electronic showing that ξ_2^{FM} is capable of localizing the occupied orbitals also for systems with a nonlocal electronic structure. However, the σ_4^{max} values indicates that the tails are thicker than for systems with a local electronic structure.

Let us now consider not only the locality of the least local orbital but of all the occupied orbitals in the set and use titin in the aug-cc-pVDZ basis as an example. The average locality of the localized occupied orbitals, σ_2^{avg} and σ_4^{avg} , are tabulated in Table 5. The maximum values are quite similar to the average values, and the largest deviation from the average values are found for the Pipek-Mezey orbitals where $\sigma_2^{\text{max}} - \sigma_2^{\text{avg}} = 1.4$ au and $\sigma_4^{\text{max}} - \sigma_4^{\text{avg}} = 1.4$ au.

In Figure 18, we have plotted the σ_2 and the σ_4 values for all occupied orbitals. The plots show a distinguishable baseline for the initial 212 orbitals which are the core orbitals. After orbital 212, the valence orbitals are plotted. The plot shows that several orbitals have σ_2 and σ_4 values close to σ_2^{max} and σ_4^{max} , and the values for those outliers are slightly larger for ξ_{PM} than for the other orbital sets. The outliers in Figure 18 represent orbitals for isolated conjugated subgroups in the titin molecule, which have a reduced flexibility for being localized as the orbitals have to describe the electron density. The exact same

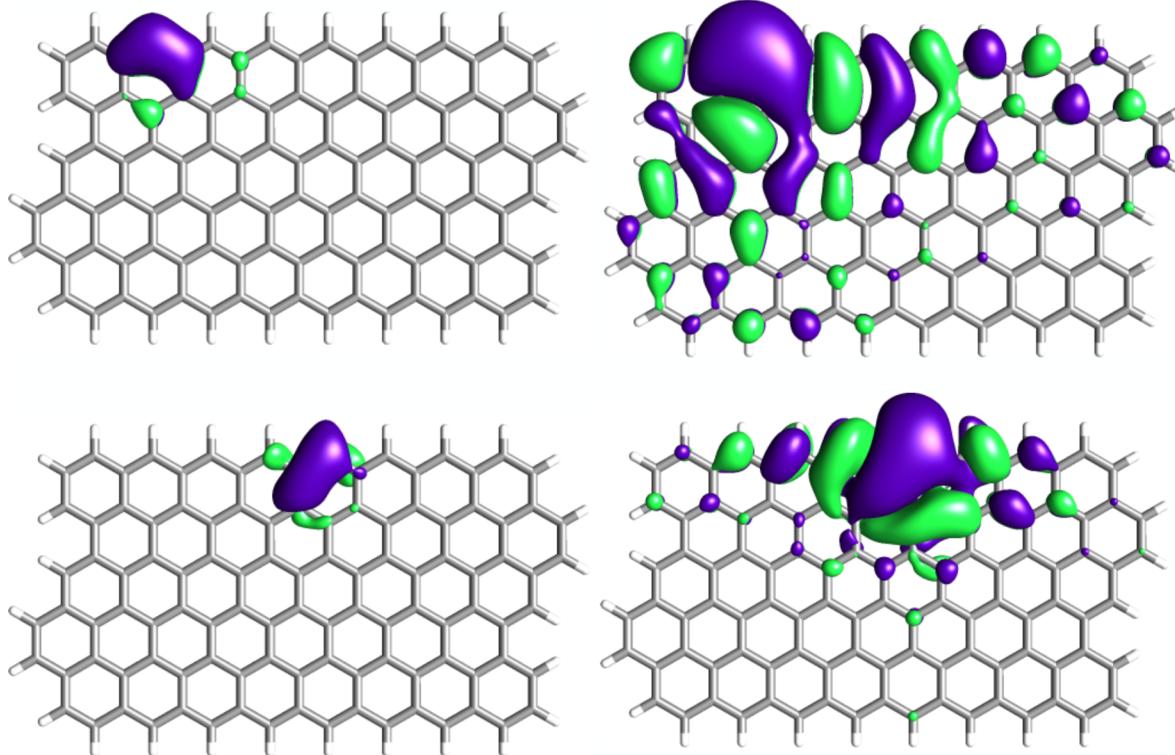


Figure 17. Least local occupied ξ_{PM} (top) and ξ_2^{FM} (bottom) orbital plotted on contour = 0.03 (left) and contour = 0.003 (right).

Table 5. σ_2^{avg} and σ_4^{avg} (in au) for Occupied Orbital Sets Obtained Using ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} for Titin in the aug-cc-pVDZ Basis Set

	σ_2^{avg}	σ_4^{avg}
ξ_{PM}	1.6	2.1
ξ_1^{SM}	1.5	2.1
ξ_2^{SM}	1.5	2.1
ξ_2^{FM}	1.5	2.1

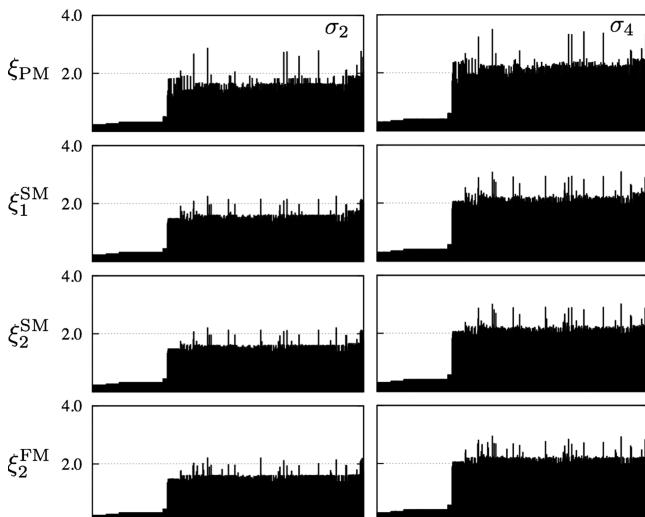


Figure 18. σ_2 and σ_4 values for all 766 occupied orbitals of titin in aug-cc-pVDZ, localized using ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} .

base lines and outliers seen in Figure 18 are also found for titin in a cc-pVDZ basis (not shown here), further emphasizing that the outliers in Figure 18 are manifestations of the physics of the

system rather than an effect of the diffuse functions in the aug-cc-pVDZ basis.

7.2. Localizing the Virtual Orbitals for Nonaugmented Basis Set

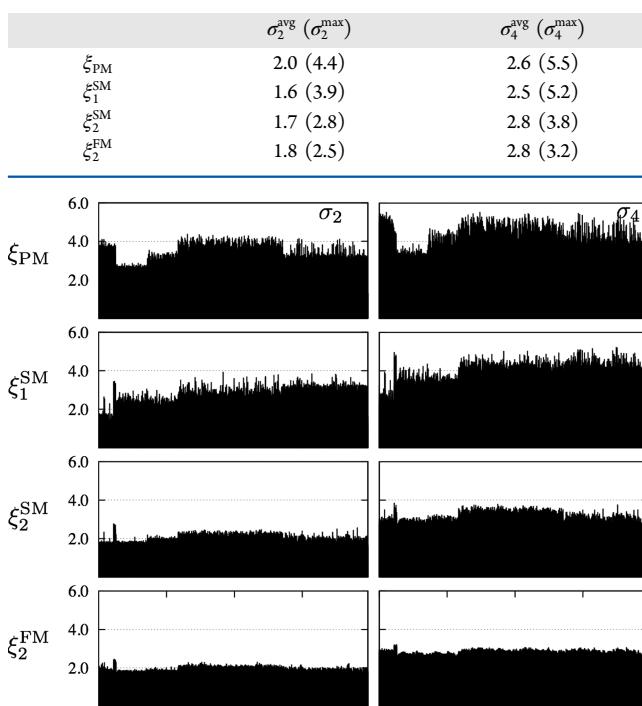
In Table 6, virtual orbital locality results, corresponding to the occupied orbital results in Table 4, are given. As seen from Table 6, the variations among the σ_2^{max} and σ_4^{max} values for the different localization schemes are larger for the virtual orbitals than for the occupied orbitals. Considering arachidic acid in the cc-pVQZ basis, we see that the difference between the smallest σ_2^{max} (for ξ_2^{FM}) and the largest σ_2^{max} (for ξ_{PM}) is 2.9 au, while the difference in σ_4^{max} values for ξ_2^{SM} and ξ_{PM} is 3.5 au. Due to the large variations in locality for the virtual orbitals, we will study the locality of the whole set of virtual orbitals in more detail and choose titin in the cc-pVTZ basis as an example.

The average values for the virtual orbital sets obtained using ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} are displayed in Table 7. Comparing the maximum values (Table 6) and the average values (Table 7), we see that the maximum values for the ξ_{PM} and ξ_1^{SM} sets are roughly twice the size of the average values, indicating the presence of outliers. In contrast, for ξ_2^{SM} and ξ_2^{FM} , the maximum and average values are more similar showing no significant outliers. To illustrate the difference in orbital locality for the different localization functions, the σ_2 and σ_4 values for all orbitals have been plotted in Figure 19. The ξ_{PM} and ξ_1^{SM} localized virtual orbitals have no baseline, while in particular the ξ_2^{FM} localized virtual orbitals have a definite baseline that is as pronounced as the baseline for the occupied orbitals in Figure 18.

To illustrate the difference in locality for the different localization functions, the orbitals which have the σ_4^{max} values are visualized in Figure 20. As seen from Figure 20, the bulk locality of the least local ξ_{PM} and ξ_1^{SM} orbitals are quite similar (as reflected by their σ_2 values) and slightly bigger than the bulk

Table 6. σ_2^{\max} and σ_4^{\max} Values (in au) for the Virtual Orbitals for Different Combinations of Molecules and Basis Sets

	ξ_{PM}		ξ_1^{SM}		ξ_2^{SM}		ξ_2^{FM}		PAO	
	σ_2^{\max}	σ_4^{\max}	σ_2^{\max}	σ_4^{\max}	σ_2^{\max}	σ_4^{\max}	σ_2^{\max}	σ_4^{\max}	σ_2^{\max}	σ_4^{\max}
aa/cc-pVDZ	3.3	4.5	3.2	4.0	2.4	3.5	2.3	3.0	3.1	3.8
aa/cc-pVTZ	4.3	5.5	3.6	4.6	2.4	3.7	2.2	3.0	3.5	3.8
aa/cc-pVQZ	5.1	6.5	3.9	5.4	2.5	3.7	2.2	3.0	3.8	4.1
aa/aug-cc-pVDZ	8.7	13.5	7.7	10.5	4.5	6.6	3.9	5.6	6.3	6.8
aa/aug-cc-pVTZ	10.4	14.6	7.2	10.3	4.7	6.4	3.7	5.2	6.8	7.3
titin/cc-pVDZ	3.7	4.9	3.5	4.3	2.7	3.6	2.5	3.2	4.7	5.6
titin/cc-pVTZ	4.4	5.5	3.9	5.2	2.8	3.8	2.5	3.2	3.7	4.5
titin/aug-cc-pVDZ	10.0	12.1	8.6	10.6	4.7	6.9	4.2	5.5	6.2	6.7
ala(20)/cc-pVTZ	4.2	5.4	3.7	4.9	2.4	3.6	2.2	2.9	3.5	3.8
ala(20)/aug-cc-pVTZ	9.8	11.5	7.6	9.4	4.6	7.0	4.1	5.3	6.5	7.0
insulin/cc-pVDZ	4.2	5.6	2.6	4.5	2.7	3.8	2.5	3.3	5.0	6.1
graphene/cc-pVDZ	4.6	6.9	3.1	4.7	2.7	4.4	2.6	3.7	6.0	10.3

Table 7. σ_2^{avg} and σ_4^{avg} (in au) for Virtual Orbital Sets Obtained Using ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} for Titin in the cc-pVTZ Basis. For Comparison, the σ_2^{\max} and σ_4^{\max} Values Are Listed in Parentheses**Figure 19.** σ_2 and σ_4 values for all 7934 virtual orbitals of titin in cc-pVTZ, localized using ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} .

of the ξ_2^{SM} and ξ_2^{FM} orbitals. The tail spread values ($\beta = 1.4$ for ξ_{PM} , ξ_1^{SM} , and ξ_2^{SM} and $\beta = 1.3$ for ξ_2^{FM}), for all four orbitals are similar. Hence, the spatial extent of the orbitals relative to each other is dictated by the magnitude of σ_2 .

Summarizing, we find that virtual orbitals of good locality can be obtained by using the more recent localization functions, ξ_2^{SM} and ξ_2^{FM} . The spatial extent of the ξ_2^{SM} and ξ_2^{FM} localized virtual orbitals is a little larger than for the occupied orbitals. However, the bulk extent for the virtual orbitals is still smaller than for the AOs. The tails of the localized orbitals are thicker for the virtual orbitals than for the occupied orbitals. Further, the locality of the virtual orbitals does not depend on whether the molecular system has a local or a nonlocal electronic structure for the ξ_2^{SM} and ξ_2^{FM} localized orbitals. For PAOs the maximum orbital

spread values are in the same range as for the ξ_1^{SM} localized orbitals. An improvement in bulk locality can therefore be obtained for the set of virtual orbitals if ξ_2^{SM} and ξ_2^{FM} localized orbitals are used.

7.3. Local Virtual Orbitals from Basis Sets Augmented with Diffuse Functions

Table 6 also contains σ_2^{\max} and σ_4^{\max} values for the localized virtual orbitals using the aug-cc-pVXZ (X = D,T) basis sets. The results for the augmented basis sets will now be discussed. The σ_2^{\max} and σ_4^{\max} are larger for augmented basis sets than for the nonaugmented basis sets. Further, the difference in locality for the localization functions is more pronounced for augmented than for nonaugmented basis sets. For example, the largest difference of σ_2^{\max} values found in Table 6 is 5.8 au (ξ_{PM} and ξ_2^{FM} , titin, aug-cc-pVDZ), and the largest difference of σ_4^{\max} values is 7.4 au (ξ_{PM} and ξ_2^{FM} , arachidic acid, aug-cc-pVTZ). As is seen from Table 6, the ξ_{PM} and ξ_1^{SM} localization functions yield large σ_2^{\max} and σ_4^{\max} values compared to the extent of atomic basis functions (see section 2.4), and the ξ_{PM} and ξ_1^{SM} orbitals can only be considered to be semilocal. The term semilocal is used since they are still much more local than CMOs.

We will use titin in the aug-cc-pVDZ basis to illustrate the locality of virtual orbitals we can obtain for augmented basis sets. The σ_2 and σ_4 values for all virtual orbitals of titin in the aug-cc-pVDZ basis are plotted in Figure 21 (note the different scale on the y-axis compared to Figure 19). Figure 21 shows that the ξ_{PM} and ξ_1^{SM} orbital sets contain so many outliers that the plots contain no apparent baseline for either σ_2 or σ_4 . For the ξ_2^{SM} and ξ_2^{FM} sets, the orbitals display similar locality and there are no outliers.

To illustrate the difference in locality for the localization functions, the least local orbitals (in terms of σ_4^{\max}) are visualized in Figure 22. From Figure 22, it is clear that the ξ_{PM} and ξ_1^{SM} orbitals show poor spatial localization. The least local ξ_2^{SM} and ξ_2^{FM} orbitals have similar σ_2 values (4.2 au and 4.1 au, respectively), while their tail spreads are different (1.6 and 1.3, respectively). As the tail spreads indicate, the least local ξ_2^{SM} orbital has a slower tail decay than the least local ξ_2^{FM} orbital. This is easily seen from the plots of the orbitals using contour = 0.001 (Figure 22, right, panels C and D). The orbitals are, however, not as local as the orbitals for the nonaugmented basis (Figure 20, panels C and D). The σ_2^{\max} and σ_4^{\max} values are found to be about two atomic units larger than the values for the nonaugmented basis set. Note that the difference in σ_2^{\max}

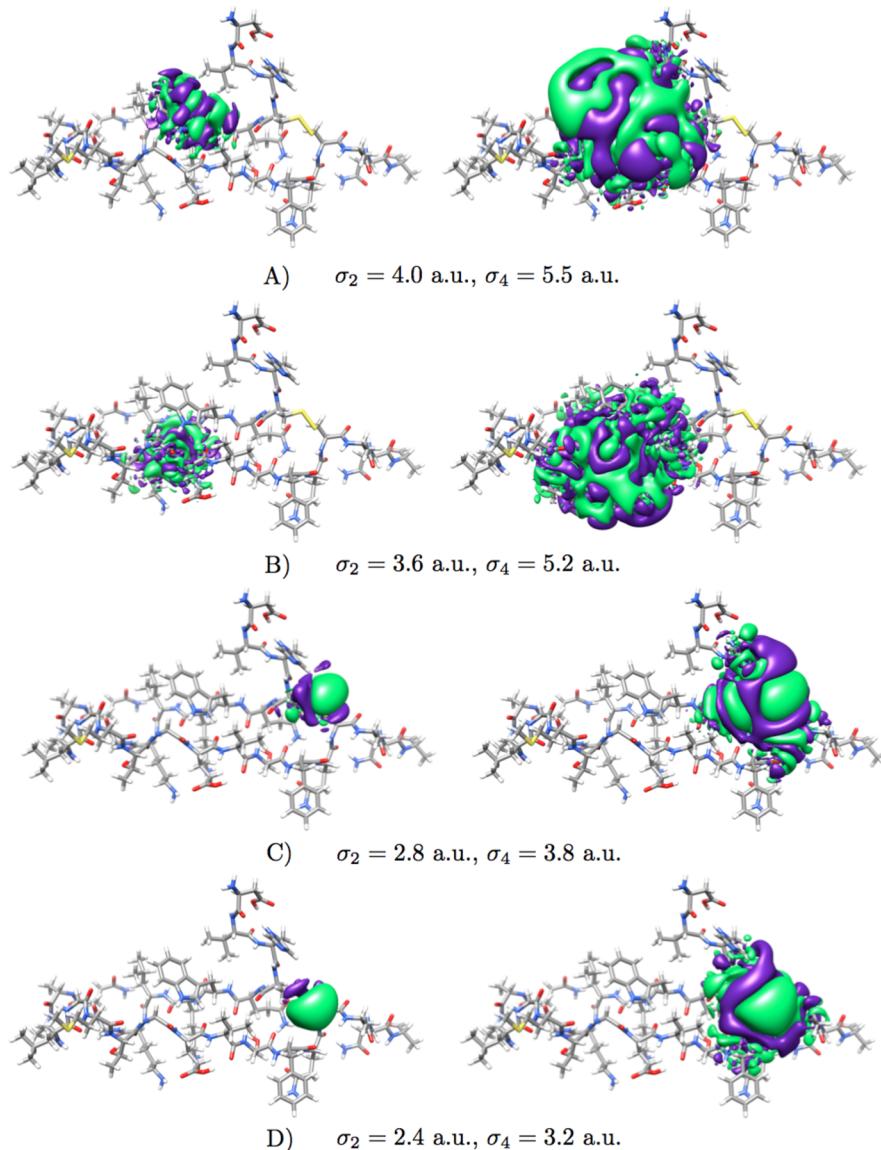


Figure 20. Least local (in terms of σ_4) virtual orbitals for titin using cc-pVTZ for (A) ξ_{PM} , (B) ξ_1^{SM} , (C) ξ_2^{SM} , and (D) ξ_2^{FM} plotted on contour = 0.01 (left) and contour = 0.001 (right).

values for the augmented and nonaugmented AOs in Table 1 is also about two atomic units.

Summarizing, we find that for Dunning's basis sets augmented with diffuse functions, only ξ_2^{SM} and ξ_2^{FM} can give a set of local virtual orbitals. Further, the improvement that is obtained using ξ_2^{FM} compared to ξ_2^{SM} is larger than for the nonaugmented basis sets.

8. EXPLORING ORBITAL LOCALITY FOR LOCAL CORRELATION METHODS

In local correlation methods, electron correlation may be described efficiently by expressing the calculations in a local Hartree–Fock basis. In a local Hartree–Fock basis, the Fock matrix elements and the integrals are local, and this locality is transferred to the wave function amplitudes when the amplitude equations are solved. It is ultimately the locality of these three quantities that makes it possible to explore the local nature of electron correlation and to obtain computationally efficient expressions for the determination of correlated wave functions for large molecular systems.

The locality of the integrals and the Fock matrix elements in the MO basis is connected to the locality of the MO charge density distributions that enter these matrix elements. The locality of the occupied-virtual charge density distribution (particle-hole excitations) is also important for an efficient determination of the solution to the Time-Dependent Hartree–Fock (TDHF) and Time-Dependent Density Functional Theory (TDDFT) eigenvalue equations as discussed by Wu et al.³⁷ We will therefore examine the locality of the charge density distributions for localized Hartree–Fock orbitals. We use numerical overlaps Ω_{PQ} (see section 5.1) as a measure of the magnitude of the charge density distributions and display the locality of charge density distributions in terms of $\max(\Omega_{PQ})$ plots.

For the alanine(20) calculation in the cc-pVTZ basis of section 4, we have in Figure 23, with a spacing of 1 au, made distance decay plots for Ω_{ij} (occupied-occupied), Ω_{ia} (occupied-virtual), and Ω_{ab} (virtual–virtual) for ξ_2^{FM} localized orbitals. The decay of the density distributions is monotonic and similar for the three distributions, with the fastest distance

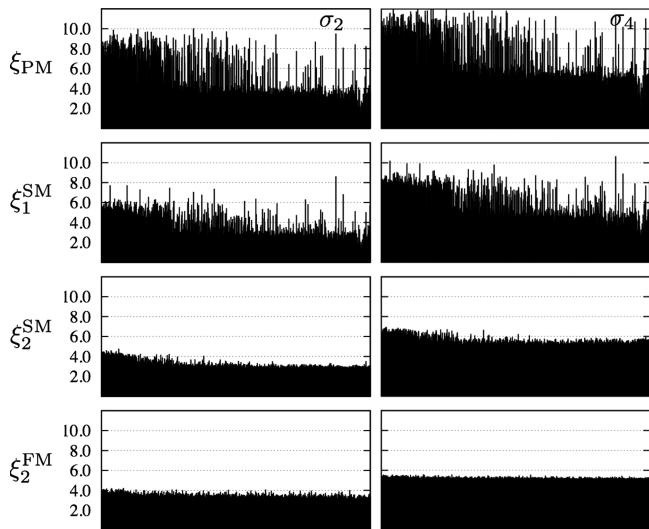


Figure 21. σ_2 and σ_4 values for all 4806 virtual orbitals of titin in aug-cc-pVDZ, localized using ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} .

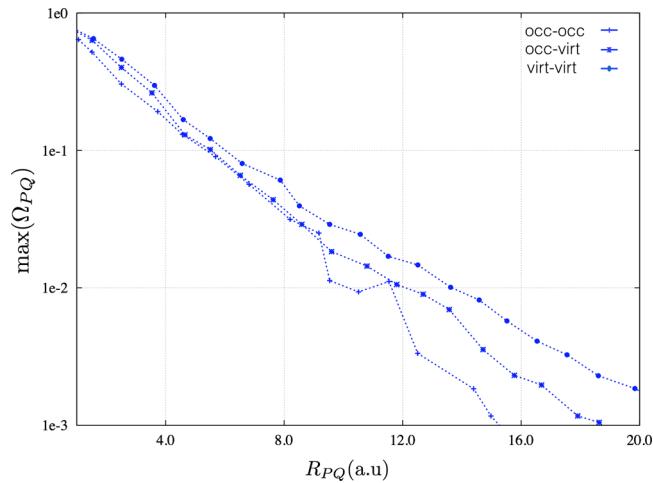


Figure 23. Maximum charge distribution for a given distance of occupied–occupied, occupied–virtual, and virtual–virtual orbitals for alanine(20) (cc-pVTZ basis) plotted for the orbital set localized using ξ_2 .

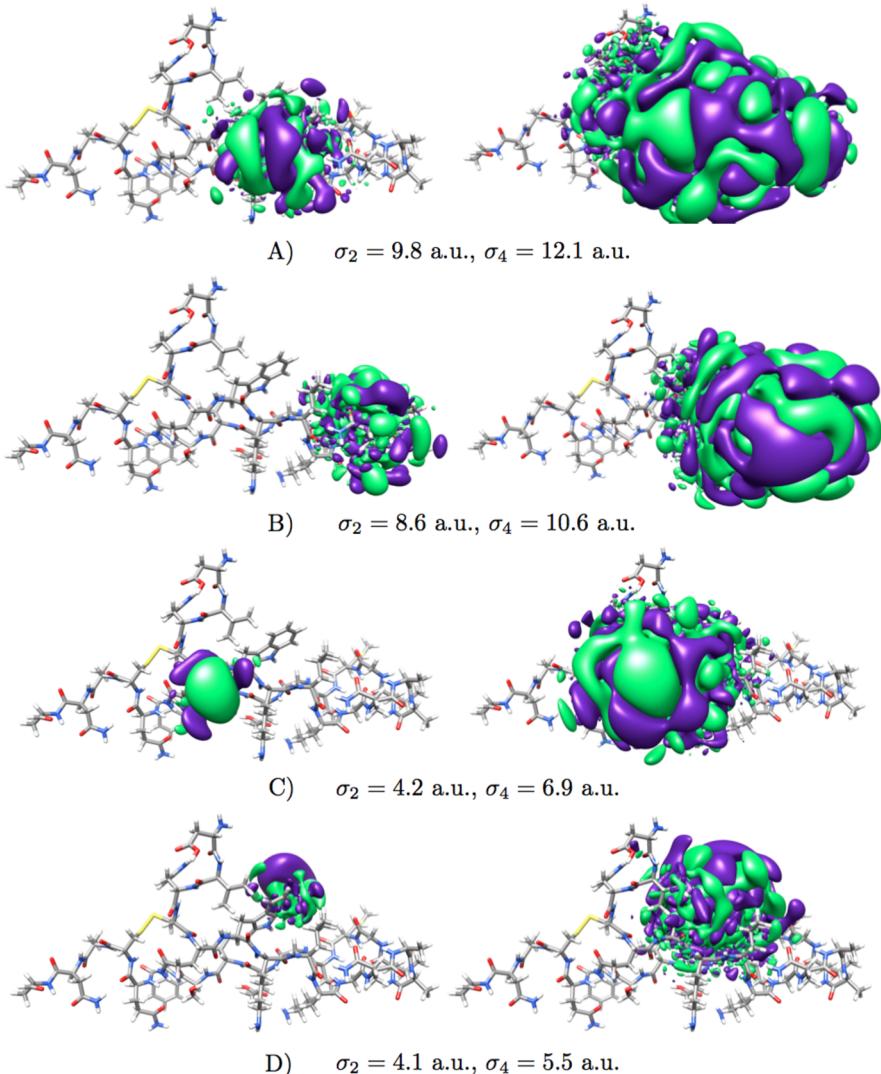


Figure 22. Least local (in terms of σ_4) virtual orbitals for titin using aug-cc-pVDZ for (A) ξ_{PM} , (B) ξ_1^{SM} , (C) ξ_2^{SM} , and (D) ξ_2^{FM} plotted on contour = 0.01 (left) and contour = 0.001 (right).

decay for the occupied–occupied distributions while the slowest is for the virtual–virtual distributions. The distance decay reflects that the orbital spread values for the occupied orbitals ($\sigma_2^{\max} = 1.6$ au) is smaller than the orbital spread for the virtual orbitals ($\sigma_2^{\max} = 2.2$ au).

The difference in orbital spreads for the ξ_2^{FM} localized occupied and virtual orbitals is small compared to the difference that is obtained for the other localization functions, in particular for ξ_1^{SM} and ξ_{PM} . To illustrate the effect this has on the distance decay of density distributions, we report for ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} density distribution plots for Ω_{ij} (Figure 24), Ω_{ia} (Figure

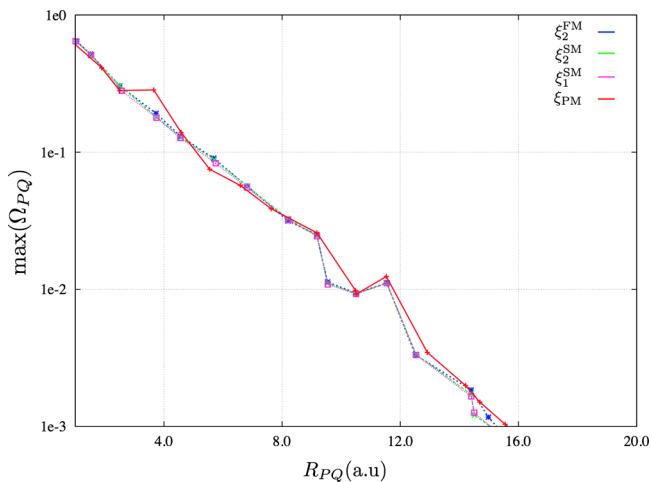


Figure 24. $\max(\Omega_{ij})$ plotted for orbital sets of alanine(20) (cc-pVTZ basis) localized using the localization functions ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} .

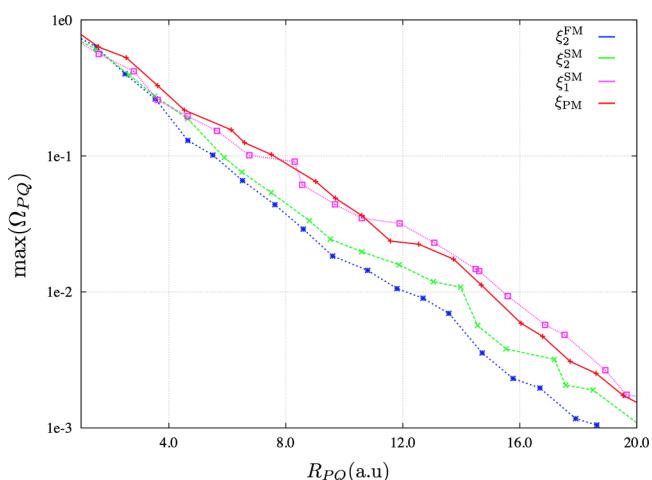


Figure 25. $\max(\Omega_{ia})$ plotted for orbital sets of alanine(20) (cc-pVTZ basis) localized using the localization functions ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} .

25), and Ω_{ab} (Figure 26). For the occupied orbitals, the orbital spreads are nearly identical for all localization functions [σ_2^{\max} is 1.9, 1.7, 1.6, and 1.6 au for ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} (see Table 4)], and consequently, the distributions in Figure 24 nearly coincide. For the virtual orbitals, the orbital spreads differ more [$(\sigma_2^{\max}$ is 3.5, 4.2, 3.7, 2.4, and 2.2 au for PAOs, ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} (see Table 6)], and this is reflected in the plots for Ω_{ia} and Ω_{ab} (Figures 25 and 26). The distance decay for Ω_{ab} is fastest for ξ_2^{FM} localized orbitals and slowest for ξ_1^{SM} and ξ_{PM}

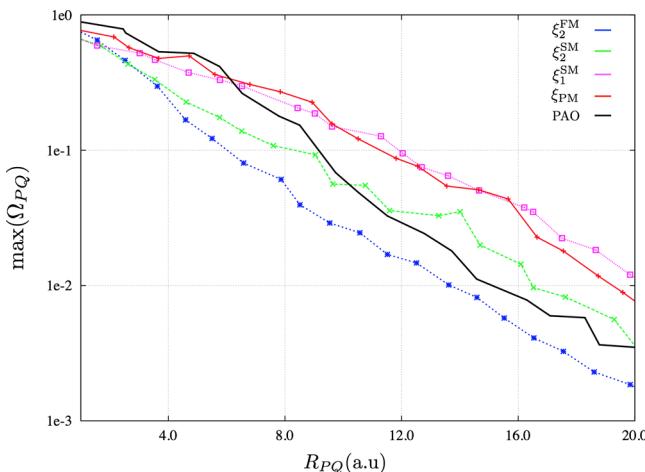


Figure 26. $\max(\Omega_{ab})$ plotted for orbital sets of alanine(20) (cc-pVTZ basis) localized using the localization functions ξ_{PM} , ξ_1^{SM} , ξ_2^{SM} , and ξ_2^{FM} . For comparison, we have also plotted the charge distribution decay for the PAOs.

localized orbitals. For the Ω_{ab} distribution plot, the distance decay for ξ_2^{SM} is between the ones for ξ_2^{FM} and the ξ_1^{SM} and ξ_{PM} localized orbitals. For distances larger than approximately 8 au, the maximum numerical overlap values are about an order of magnitude smaller for the ξ_2^{SM} localized orbitals than for the ξ_1^{SM} and ξ_{PM} localized orbitals. The distance decay for the PAOs is similar to the ones for ξ_{PM} and ξ_1^{SM} up to 7 au, but thereafter, the decay is more rapid and ends up between the ξ_2^{SM} and ξ_2^{FM} localized orbitals.

Summarizing, we find a close connection between the size of the maximum orbital spreads and the distance decay of the density distributions (numerical overlaps). This connection results in a faster distance decay of the virtual–virtual density distributions for ξ_2^{FM} localized orbitals than for the orbitals that are localized using the other localization functions, while the occupied–occupied density distribution decays are essentially the same for orbitals generated by the different localization functions.

9. CONCLUDING REMARKS

The spatial localization of occupied and virtual Hartree–Fock orbitals has been the subject of many investigations for several decades, due to their conceptual and computational importance in electronic structure theory. In this review, we have presented an overview of the developments with respect to the characterization and generation of spatially localized orthogonal Hartree–Fock orbitals. In particular, the developments have been seen in the context of the recent advances that has been made in optimization of orbital localization functions. In order to carry out a detailed analysis of the spatial locality of local orbitals generated using traditional and more modern localization functions, we have provided an in-depth analysis on the characterization of locality based on measures which relate both the orbital bulk and the orbital tail behavior. The discussion of locality is made more lucid by the use of a reference locality, the atomic orbitals, and complemented by considerations on the localizability of the orbitals. With regard to the localizability of the orbitals, the review has also emphasized that the failure to generate local orbitals for electronically nonlocal systems has been due to inadequacies in the used localization schemes, rather than to the inherent nonlocal electronic structure of the

molecular system. The developments summarized in the present review show that occupied and virtual Hartree–Fock orbitals of good locality can be generated for both electronically local and nonlocal molecular systems and when using diffuse basis functions. For localization beyond the localization of occupied orbitals of electronically local systems, localization functions will yield orbitals of varying locality, and the recently introduced localization functions are seen to be superior to the traditional ones.

AUTHOR INFORMATION

Corresponding Author

*E-mail: [ida-marie.hoyvik@ntnu.no](mailto:idamarie.hoyvik@ntnu.no).

Notes

The authors declare no competing financial interest.

Biographies

Ida-Marie Høyvik is a fixed-term associate professor at the Center for Applied Theoretical Chemistry at the Norwegian University of Science and Technology (Norway). She received her Ph.D. degree from Aarhus University (Denmark) in 2013 and was a postdoctoral researcher at Ludwig-Maximilians-Universität Munich (Germany). Her research topics are local coupled-cluster theory and localization of orbitals.

Poul Jørgensen is a Professor at Aarhus University (Denmark), where he received his doctoral degree in 1976. He is a coauthor of more than 300 papers and 3 monographs. His interests consist of determination of molecular properties using response function theory and the description of large molecular systems using local coupled-cluster theory.

ACKNOWLEDGMENTS

P.J. acknowledges support from The European Research Council under the European Union's (EU) Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 291371 and support from the Danish Council of Independent Research/Natural Sciences.

REFERENCES

- (1) Lennard-Jones, J. E. The Molecular Orbital Theory of Chemical Valency I. The Determination of Molecular Orbitals. *Proc. R. Soc. London, Ser. A* **1949**, *198*, 1–13.
- (2) Lennard-Jones, J. E. The Molecular Orbital Theory of Chemical Valency II. Equivalent Orbitals in Molecules of Known Symmetry. *Proc. R. Soc. London, Ser. A* **1949**, *198*, 14–26.
- (3) Hall, G.; Lennard-Jones, J. E. The Molecular Orbital Theory of Chemical Valency III. Properties of Molecular Orbitals. *Proc. R. Soc. London, Ser. A* **1950**, *202*, 155–165.
- (4) Lennard-Jones, J. E.; Pople, J. The Molecular Orbital Theory of Chemical Valency IV. The Significance of Equivalent Orbitals. *Proc. R. Soc. London, Ser. A* **1950**, *202*, 166–180.
- (5) Hall, G. The Molecular Orbital Theory of Chemical Valency VI. Properties of Equivalent Orbitals. *Proc. R. Soc. London, Ser. A* **1950**, *202*, 336–344.
- (6) Boys, S. F. Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.* **1960**, *32*, 296–299.
- (7) Boys, S. F. *Quantum Theory of Atoms, Molecules and Solid State*; Academic: New York, 1966; p 253.
- (8) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457–464.
- (9) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. II. *J. Chem. Phys.* **1965**, *43*, S97–S116.
- (10) Magnasco, V.; Perico, A. Uniform Localization of Atomic and Molecular Orbitals. I. *J. Chem. Phys.* **1967**, *47*, 971–981.
- (11) Magnasco, V.; Perico, A. Uniform Localization of Atomic and Molecular Orbitals. II. *J. Chem. Phys.* **1968**, *48*, 800–808.
- (12) Caldwell, D.; Redington, P.; Eyring, H. Density Matrix Method for Orbital Localization. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 3042–3045.
- (13) Foster, J. P.; Weinhold, F. Natural Hybrid Orbitals. *J. Am. Chem. Soc.* **1980**, *102*, 7211–7218.
- (14) Reed, A. E.; Weinhold, F. Natural Localized Molecular Orbitals. *J. Chem. Phys.* **1985**, *83*, 1736–1740.
- (15) Pipek, J.; Mezey, P. G. A Fast Intrinsic Localization Procedure Applicable for Ab Initio and Semiempirical Linear Combination of Atomic Orbital Wave Functions. *J. Chem. Phys.* **1989**, *90*, 4916–4925.
- (16) von Niessen, W. Density Localization of Atomic and Molecular Orbitals. I. *J. Chem. Phys.* **1972**, *56*, 4290–4297.
- (17) Bartha, F.; Bogar, F.; Kapuy, E. Localization of Virtual Orbitals. *Int. J. Quantum Chem.* **1990**, *38*, 215–219.
- (18) Cioslowski, J. Partitioning of the Orbital Overlap Matrix and the Localization Criteria. *J. Math. Chem.* **1991**, *8*, 169–178.
- (19) Rubio, J.; Povill, A.; Malrieu, J.; Reinhardt, P. Direct Determination of Localized Hartree-Fock Orbitals as a Set Toward N Scaling Procedures. *J. Chem. Phys.* **1997**, *107*, 10044–10050.
- (20) Liu, S.; Perez-Jorda, J.; Yang, W. Nonorthogonal Localized Molecular Orbitals in Electronic Structure Theory. *J. Chem. Phys.* **2000**, *112*, 1634–1644.
- (21) Gu, F. L.; Aoki, Y.; Korchowiec, J.; Imamura, A.; Kirtman, B. A New Localization Scheme for the Elongation Method. *J. Chem. Phys.* **2004**, *121*, 10385–10391.
- (22) Feng, H.; Bian, J.; Li, L.; Yang, W. An Efficient Method for Constructing Nonorthogonal Localized Molecular Orbitals. *J. Chem. Phys.* **2004**, *120*, 9458–9466.
- (23) Subotnik, J. E.; Shao, Y.; Liang, W.; Head-Gordon, M. An Efficient Method for Calculating Maxima of Homogeneous Functions of Orthogonal Matrices: Applications to Localized Occupied Orbitals. *J. Chem. Phys.* **2004**, *121*, 9220–9229.
- (24) Subotnik, J. E.; Dutoi, A. D.; Head-Gordon, M. Fast Localized Orthonormal Virtual Orbitals Which Depend Smoothly on Nuclear Coordinates. *J. Chem. Phys.* **2005**, *123*, 114108–1–114108–9.
- (25) Subotnik, J. E.; Head-Gordon, M. A Local Correlation Model That Yields Intrinsically Smooth Potential-Energy Surfaces. *J. Chem. Phys.* **2005**, *123*, 064108–1–064108–9.
- (26) Alcoba, D. R.; Lain, L.; Torre, A.; Bochicchio, R. C. An Orbital Localization Criterion Based on the Theory of "Fuzzy" Atoms. *J. Comput. Chem.* **2006**, *27*, 596–608.
- (27) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. Localized Orbital Theory and Ammonia Triborane. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5522–5530.
- (28) Aquilante, F.; Pedersen, T. B.; de Meras, A. S.; Koch, H. Fast Noniterative Orbital Localization for Large Molecules. *J. Chem. Phys.* **2006**, *125*, 174101–1–174101–7.
- (29) Weber, V.; Hutter, J. A Smooth $\ell(1)$ -Norm Sparseness Function for Orbital Based Linear Scaling Total Energy Minimization. *J. Chem. Phys.* **2008**, *128*, 064107–1–064107–6.
- (30) Ziolkowski, M.; Jansik, B.; Jørgensen, P.; Olsen, J. Maximum Locality in Occupied and Virtual Spaces Using a Least-Change Strategy. *J. Chem. Phys.* **2009**, *131*, 124112–1–124112–15.
- (31) Cui, G.; Fang, W.; Yang, W. Efficient Construction of Nonorthogonal Localized Molecular Orbitals in Large Systems. *J. Phys. Chem. A* **2010**, *114*, 8878–8883.
- (32) Jansik, B.; Høst, S.; Kristensen, K.; Jørgensen, P. Local Orbitals by Minimizing Powers of the Orbital Variance. *J. Chem. Phys.* **2011**, *134*, 194104–1–194104–9.
- (33) Zoboki, T.; Mayer, I. Extremely Localized Nonorthogonal Orbitals by the Pairing Theorem. *J. Comput. Chem.* **2011**, *32*, 689–695.
- (34) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Orbital Localization Using Fourth Central Moment Minimization. *J. Chem. Phys.* **2012**, *137*, 224114–1–224114–11.

- (35) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Pipek-Mezey Localization of Occupied and Virtual Orbitals. *J. Comput. Chem.* **2013**, *34*, 1456–1462.
- (36) de Silva, P.; Giebultowski, M.; Korchowiec, J. Fast Orbital Localization Scheme in Molecular Fragments Resolution. *Phys. Chem. Chem. Phys.* **2012**, *14*, 546–552.
- (37) Wu, F.; Liu, W.; Zhang, Y.; Li, Z. Linear-Scaling Time-Dependent Density Functional Theory Based on the Idea of "From Fragments to Molecule. *J. Chem. Theory Comput.* **2011**, *7*, 3643–3660.
- (38) Li, Z.; Li, H.; Suo, B.; Liu, W. Localization of Molecular Orbitals: From Fragments to Molecule. *Acc. Chem. Res.* **2014**, *47*, 2758–2767.
- (39) Zhang, C.; Li, S. An Efficient Localization Procedure for Large Systems Using a Sequential Transformation Strategy. *J. Chem. Phys.* **2014**, *141*, 244106–1–244106–8.
- (40) Zimmerman, P. M.; Molina, A. R.; Smereka, P. Orbitals With Intermediate Localization and Low Coupling: Spanning the Gap Between Canonical and Localized Orbitals. *J. Chem. Phys.* **2015**, *143*, 014106–1–014106–8.
- (41) Marynick, D. S.; Lipscomb, W. N. Self-Consistent Field and Localized Orbital Study of 4,5-dicarbahexaborane(8). *J. Am. Chem. Soc.* **1972**, *94*, 8699–8706.
- (42) Wolfe, S.; Tel, L. M.; Haines, W. J.; Robb, M. A.; Csizmadia, I. G. Gauche effect. Localized Molecular Orbitals and Excited-State Geometries in Fluoromethanol. *J. Am. Chem. Soc.* **1973**, *95*, 4863–4870.
- (43) Brown, L. D.; Kleier, D. A.; Lipscomb, W. N. Localized Molecular Orbitals for Carbon Dioxide and Carbonate(2-) Ion. A Comparison of Localization Types and a Comment on Isoelectronic Structures. *J. Am. Chem. Soc.* **1977**, *99*, 6787–6792.
- (44) Subotnik, J. E.; Vura-Weis, J.; Sodt, A. J.; Ratner, M. A. Predicting Accurate Electronic Excitation Transfer Rates via Marcus Theory with Boys or Edmiston-Ruedenberg Localized Diabatization. *J. Phys. Chem. A* **2010**, *114*, 8665–8675.
- (45) Pavanello, M.; Neugebauer, J. Linking the Historical and Chemical Definitions of Diabatic States for Charge and Excitation Energy Transfer Reactions in Condensed Phase. *J. Chem. Phys.* **2011**, *135*, 134113–1–134113–6.
- (46) McAllister, L. J.; Bruce, D. W.; Karadakov, P. B. Quantum Chemical Investigation of Attractive Non-Covalent Interactions between Halomethanes and Rare Gases. *J. Phys. Chem. A* **2012**, *116*, 10621–10628.
- (47) Dubillard, S.; Rota, J.; Saue, T.; Fægri, K. Bonding Analysis Using Localized Relativistic Orbitals: Water, the Ultrarelativistic Case and the Heavy Homologues H₂X. *J. Chem. Phys.* **2006**, *124*, 154307–1–154307–14.
- (48) Standara, S.; Bouzkova, K.; Straka, M.; Zacharova, Z.; Hocek, M.; Marek, J.; Marek, R. Interpretation of Substituent Effects on C-13 and N-15 NMR Chemical Shifts in 6-substituted Purines. *Phys. Chem. Chem. Phys.* **2011**, *13*, 15854–15864.
- (49) West, A. C.; Schmidt, M. W.; Gordon, M. S.; Ruedenberg, K. A Comprehensive Analysis of Molecule-Intrinsic Quasi-Atomic, Bonding, and Correlating Orbitals. I. Hartree-Fock Wave Functions. *J. Chem. Phys.* **2013**, *139*, 234107–1–234107–19.
- (50) Sinanoglu, O. Many-Electron Theory of Atoms, Molecules and Their Interactions. *Advances in Chemical Physics; Advances in Chemical Physics*; John Wiley & Sons: Hoboken, NJ, 1964; Vol. 6; pp 315–41210.1002/9780470143520.ch7.
- (51) Nesbet, R. Electronic Correlation in Atoms and Molecules. *Adv. Chem. Phys.* **1965**, *9*, 321.
- (52) Diner, S.; Malrieu, J.; Claverie, P.; Jordan, F. Fully Localized Bond Orbitals and the Correlation Problem. *Chem. Phys. Lett.* **1968**, *2*, 319–323.
- (53) Pulay, P. Localizability of Dynamic Electron Correlation. *Chem. Phys. Lett.* **1983**, *100*, 151–154.
- (54) Sæbø, S.; Pulay, P. Local Configuration-Interaction - and Efficient Approach for Larger Molecules. *Chem. Phys. Lett.* **1985**, *113*, 13–18.
- (55) Hampel, C.; Werner, H.-J. Local Treatment of Electron Correlation in Coupled Cluster Theory. *J. Chem. Phys.* **1996**, *104*, 6286–6297.
- (56) Christiansen, O.; Manninen, P.; Jørgensen, P.; Olsen, J. Coupled-Cluster Theory in a Projected Orbital Basis. *J. Chem. Phys.* **2006**, *124*, 084103–1–084103–9.
- (57) Löwdin, P. O. Quantum Theory of Many-Particle Systems. I. Physical Interpretations by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configurational Interaction. *Phys. Rev.* **1955**, *97*, 1474–1489.
- (58) Neese, F.; Wennmohs, F.; Hansen, A. Efficient and Accurate Local Approximations to Coupled-Electron Pair Approaches: An Attempt to Revive the Pair Natural Orbital Method. *J. Chem. Phys.* **2009**, *130*, 1141108–1–1141108–18.
- (59) Liakos, D. G.; Neese, F. Domain Based Pair Natural Orbital Coupled Cluster Studies on Linear and Folded Alkane Chains. *J. Chem. Theory Comput.* **2015**, *11*, 2137–2143.
- (60) Yang, J.; Kurashige, Y.; Manby, F. R.; Chan, G. K. L. Tensor Factorizations of Local Second-Order Møller-Plesset Theory. *J. Chem. Phys.* **2011**, *134*, 044123–1–044123–13.
- (61) Yang, J.; Chan, G. K. L.; Manby, F. R.; Schütz, M.; Werner, H. J. The orbital-Specific-Virtual Local Coupled Cluster Singles and Doubles method. *J. Chem. Phys.* **2012**, *136*, 144105–1–144105–16.
- (62) Ziolkowski, M.; Jansik, B.; Kjærgaard, T.; Jørgensen, P. Linear Scaling Coupled Cluster Method With Correlation Energy Based Error Control. *J. Chem. Phys.* **2010**, *133*, 014107–1–014107–5.
- (63) Eriksen, J.; Baudin, P.; Ettenhuber, P.; Kristensen, K.; Kjærgaard, T.; Jørgensen, P. Linear-Scaling Coupled Cluster with Perturbative Triple Excitations: the Divide-Expand-Consolidate CCSD(T) Model. *J. Chem. Theory Comput.* **2015**, *11*, 2984–2993.
- (64) Cullen, J. M.; Zerner, M. C. The Linked Singles and Doubles Model: An Approximate Theory of Electron Correlation Based on the Coupled-Cluster Ansatz. *J. Chem. Phys.* **1982**, *77*, 4088–4109.
- (65) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. Electron Correlation Theories and Their Application To the Study of Simple Reaction Potential Surfaces. *Int. J. Quantum Chem.* **1978**, *14*, 545–560.
- (66) Laidig, W. D.; Purvis, G. D.; Bartlett, R. J. Localized orbitals in the Coupled Cluster Singles and Doubles Model. *Int. J. Quantum Chem.* **1982**, *22*, 561–573.
- (67) Förner, W.; Ladik, J.; Otto, P.; Cizek, J. Coupled-cluster studies. II. The Role of Localization in Correlation Calculations on Extended Systems. *Chem. Phys.* **1985**, *97*, 251–262.
- (68) Takahashi, M.; Paldus, J. Coupled-Cluster Approach to Electron Correlation in One Dimension. II. Cyclic Polyene Model in Localized Basis. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1985**, *31*, 5121–5142.
- (69) Sæbø, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213–236.
- (70) Schütz, M.; Hetzer, G.; Werner, H.-J. Low-Order Scaling Local Electron Correlation Methods. I. Linear scaling Local MP2. *J. Chem. Phys.* **1999**, *111*, 5691–5705.
- (71) Flocke, N.; Bartlett, R. J. A Natural Linear Scaling Coupled-Cluster Method. *J. Chem. Phys.* **2004**, *121*, 10935–10944.
- (72) Li, S.; Ma, J.; Jiang, Y. Linear Scaling Local Correlation Approach for Solving the Coupled Cluster Equations of Large Systems. *J. Comput. Chem.* **2002**, *23*, 237–244.
- (73) Kobayashi, M.; Nakai, H. Extension of Linear-Scaling Divide-and-Conquer-Based Correlation Method to Coupled Cluster Theory with Singles and Doubles Excitations. *J. Chem. Phys.* **2008**, *129*, 044103–1–044103–9.
- (74) Fedorov, D. G.; Kitaura, K. Coupled-Cluster Theory Based Upon the Fragment Molecular-Orbital Method. *J. Chem. Phys.* **2005**, *123*, 134103–1–134103–11.
- (75) Stoll, H. The Correlation Energy of Crystalline Silicon. *Chem. Phys. Lett.* **1992**, *191*, 548–552.
- (76) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. A Near Linear-Scaling Smooth Local Coupled Cluster Algorithm for Electronic Structure. *J. Chem. Phys.* **2006**, *125*, 074116–1–074116–12.
- (77) Maslen, P. E.; Lee, M. S.; Head-Gordon, M. An Accurate Local Model for Triple Substitutions in Fourth Order Møller-Plesset Theory

- and in Perturbative Corrections to Singles and Doubles Coupled Cluster Methods. *Chem. Phys. Lett.* **2000**, *319*, 205–212.
- (78) Kapuy, E.; Bartha, F.; Bogar, F.; Csepé, Z.; Kozmutza, C. Applications of the MBPT in the Localized Representation. *Int. J. Quantum Chem.* **1990**, *38*, 139–147.
- (79) Ayala, P. Y.; Scuseria, G. E. Linear Scaling Second-Order Møller-Plesset Theory in the Atomic Orbital Basis for Large Molecular Systems. *J. Chem. Phys.* **1999**, *110*, 3660–3671.
- (80) Lambrecht, D. S.; Doser, B.; Ochsenfeld, C. Rigorous Integral Screening for Electron Correlation Methods. *J. Chem. Phys.* **2005**, *123*, 184102–1–184102–11.
- (81) Myhre, R. H.; Sanchez de Meras, A.; Koch, H. Multi-Level Coupled Cluster Theory. *J. Chem. Phys.* **2014**, *141*, 224105–1–224105–11.
- (82) Finley, J. P.; Hirao, K. Multireference Moller-Plesset Perturbation Theory with Non-Canonical and Non-Orthogonal Orbitals. *Chem. Phys. Lett.* **2000**, *328*, 60–66.
- (83) Angeli, C.; Calzado, C. J.; Cimiraglia, R.; Evangelisti, S.; Guihéry, N.; Leininger, T.; Malrieu, J.-P.; Maynau, D.; Ruiz, J. V. P.; Sparta, M. The Use of Local Orbitals in Multireference Calculations. *Mol. Phys.* **2003**, *101*, 1389–1398.
- (84) Chwee, T. S.; Carter, E. A. Cholesky Decomposition Within Local Multireference Singles and Doubles Configuration Interaction. *J. Chem. Phys.* **2010**, *132*, 074104–1–074104–10.
- (85) Ben Amor, N.; Bessac, F.; Hoyau, S.; Maynau, D. Direct Selected Multireference Configuration Interaction Calculations for Large Systems Using Localized Orbitals. *J. Chem. Phys.* **2011**, *135*, 014101–1–014101–14.
- (86) Chang, C.; Calzado, C. J.; Ben Amor, N.; Sanchez Marin, J.; Maynau, D. Multi-Scale Multireference Configuration Interaction Calculations for Large Systems Using Localized Orbitals: Partition in Zones. *J. Chem. Phys.* **2012**, *137*, 104102–1–104102–12.
- (87) Beran, G.; Hirata, S. Fragment and Localized Orbital Methods in Electronic Structure Theory. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7559–7561.
- (88) Crawford, T. D. Reduced-Scaling Coupled-Cluster Theory for Response Properties of Large Molecules In *Recent Progress in Coupled Cluster Methods; Challenges and Advances in Computational Chemistry and Physics*; Carsky, P., Paldus, J., Pittner, J., Eds.; Springer: Netherlands, 2010; Vol. 11; pp 37–55.
- (89) Lehtola, S.; Jonsson, H. Pipek-Mezey Orbital Localization Using Various Partial Charge Estimates. *J. Chem. Theory Comput.* **2014**, *10*, 642–649.
- (90) Burke, L.; Leroy, G.; Daudel, R.; Stephens, M. Absolute Overlap and Second Moment Dispersions as Measures of Localizability in the Spatial and Energetic Methods. *Chem. Phys. Lett.* **1978**, *57*, 15–21.
- (91) Berghold, G.; Mundy, C. J.; Romero, A. H.; Hutter, J.; Parrinello, M. General and Efficient Algorithms for Obtaining Maximally Localized Wannier Functions. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2000**, *61*, 10040–10048.
- (92) Pipek, J. Localization Measure and Maximum Delocalization in Molecular Systems. *Int. J. Quantum Chem.* **1989**, *36*, 487–501.
- (93) Fornili, A.; Sironi, M.; Raimondi, M. Determination of Extremely Localized Molecular Orbitals and Their Application to Quantum Mechanics Molecular Mechanics Methods and to the Study of Intramolecular Hydrogen Bonding. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 157–172.
- (94) Knizia, G. Intrinsic Atomic Orbitals: An Unbiased Bridge Between Quantum Theory and Chemical Concepts. *J. Chem. Theory Comput.* **2013**, *9*, 4834–4843.
- (95) Kenney, J. F. *Mathematics of Statistics, Part I*; Chapman & Hall LTD., 1939; Chapter IV.
- (96) Millie, P.; Levy, B.; Berthier, G. *Localization and Delocalization in Quantum Chemistry*; Springer: Netherlands, 1975; Vol. 1; pp 59–97.
- (97) Daudel, R. *Localization and Delocalization in Quantum Chemistry*; Springer: Netherlands, 1975; Vol. 1; pp 111–112.
- (98) Malrieu, J. *Localization and Delocalization in Quantum Chemistry*; Springer: Netherlands, 1975; Vol. 1; pp 249–249.
- (99) Høyvik, I.-M.; Kristensen, K.; Kjærgaard, T.; Jørgensen, P. A Perspective on the Localizability of Hartree-Fock Orbitals. *Theor. Chem. Acc.* **2014**, *133* (1417), 1–10.
- (100) Taylor, W. J. Localized Molecular Orbitals by a Quadratic Approximation Method of Steepest Ascents. *J. Chem. Phys.* **1968**, *48*, 2385–2386.
- (101) Kari, R. *Progress in Theoretical Organic Chemistry*; Elsevier, 1977.
- (102) Ryback, W.; Poirier, R.; Kari, R. Application of Method of Conjugate Gradients to Calculation of Minimal Basis Set Localized Orbitals. *Int. J. Quantum Chem.* **1978**, *13*, 1–16.
- (103) Lehtola, S.; Jonsson, H. Unitary Optimization of Localized Molecular Orbitals. *J. Chem. Theory Comput.* **2013**, *9*, 5365–5372.
- (104) Leonard, J. M.; Luken, W. L. Quadratically Convergent Calculation of Localized Molecular Orbitals. *Theor. Chim. Acta* **1982**, *62*, 107–132.
- (105) Kari, R. Parametrization and Comparative-Analysis of the BFGS Optimization Algorithm for the Determination of Optimum Linear Coefficients. *Int. J. Quantum Chem.* **1984**, *25*, 321–329.
- (106) Fletcher, R. *Practical Methods of Optimization*, 2nd Ed.; Wiley: Chichester, 1987.
- (107) Davidson, E. R. Iterative Calculation of a Few of Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-Symmetric Matrices. *J. Comput. Phys.* **1975**, *17*, 87–94.
- (108) Høyvik, I.-M.; Jørgensen, P. Localized Orbitals From Basis Sets Augmented with Diffuse Functions. *J. Chem. Phys.* **2013**, *138*, 204104–1–204104–7.
- (109) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Trust-Region Minimization for Orbital Localization Functions. *J. Chem. Theory Comput.* **2012**, *8*, 3137–3146.
- (110) Cederbaum, L.; Schirmer, J.; Meyer, H.-D. Block Diagonalization of Hermitian Matrices. *J. Phys. A: Math. Gen.* **1989**, *22*, 2427–2439.
- (111) Guo, Y.; Li, W.; Li, S. An Efficient Linear Scaling Procedure for Constructing Localized Orbitals of Large Molecules Based on the One-Particle Density Matrix. *J. Chem. Phys.* **2011**, *135*, 134107–1–134107–7.
- (112) Høyvik, I.-M.; Jansik, B.; Kristensen, K.; Jørgensen, P. Local Hartree-Fock Orbitals Using a Three-Level Optimization Strategy for the Energy. *J. Comput. Chem.* **2013**, *34*, 1311–1320.
- (113) Marzari, N.; Vanderbilt, D. Maximally Localized Generalized Wannier Functions for Composite Energy Bands. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *56*, 12847–12865.
- (114) Darlington, R. B. Is Kurtosis Really ‘Peakedness’? *Am. Stat.* **1970**, *24*, 19–22.
- (115) DeCarlo, L. T. On the Meaning and Use of Kurtosis. *Psychological methods* **1997**, *2*, 292–307.
- (116) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (117) Kendall, R.; Dunning, T.; Harrison, R. Electron Affinities of the 1st Row Atoms Revisited - Systematic Basis-Sets and Wave-Functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (118) Woon, D. E.; Dunning, T. Gaussian Basis Sets for Use in Correlated Molecular Calculations. 2. The Atoms Aluminum Through Argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (119) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. *J. Chem. Phys.* **1955**, *23*, 1833–1840.
- (120) Weinhold, F. *Discovering Chemistry With Natural Bond Orbitals*; Wiley, 2012; Chapter 3.2.
- (121) Mulliken, R. S.; Ermel, W. C. *Diatomic Molecules, Results of Ab Initio Calculations*; Academic Press, 1977; pp 33–38.
- (122) Löwdin, P.-O. On the Non-Orthogonality Problem Connected With the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365–375.
- (123) Baba, T.; Takeuchi, M.; Nakai, H. Natural Atomic Orbital Based Energy Density Analysis: Implementation and Applications. *Chem. Phys. Lett.* **2006**, *424*, 193–198.

- (124) Kern, C. W.; Karplus, M. Analysis of Charge Distributions: Hydrogen Fluoride. *J. Chem. Phys.* **1964**, *40*, 1374–1389.
- (125) Politzer, P.; Mulliken, R. S. Comparison of 2 Atomic Charge Definitions, as Applied to Hydrogen Fluoride Molecule. *J. Chem. Phys.* **1971**, *55*, 5135–5136.
- (126) Ruedenberg, K. *Localization and Delocalization in Quantum Chemistry*; Springer: Netherlands, 1975; Vol. 1; pp 223–245.
- (127) Kleier, D. A.; Halgren, T. A.; Hall, J. H.; Lipscomb, W. N. Localized Molecular Orbitals for Polyatomic Molecules. I. A Comparison of the Edmiston-Ruedenberg and Boys Localization Methods. *J. Chem. Phys.* **1974**, *61*, 3905–3919.
- (128) Kleier, D. A.; Hall, J. H.; Halgren, T. A.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 2265–2267.
- (129) Kozmutza, C.; Kapuy, E. Localized Orbitals for the Description of Molecular Interaction. *Int. J. Quantum Chem.* **1990**, *38*, 665–673.
- (130) Kapuy, E.; Szalay, P.; Daudel, R.; Bogar, F.; Bartha, F.; Kozmutza, C. The Transferability of Some Molecular Properties in All-Trans Conjugated Polyenes. *J. Mol. Struct.: THEOCHEM* **1991**, *226*, 351–355.
- (131) Philipp, D.; Friesner, R. Mixed Ab Initio QM/MM Modeling Using Frozen Orbitals and Tests With Alanine Dipeptide and Tetrapeptide. *J. Comput. Chem.* **1999**, *20*, 1468–1494.
- (132) Autschbach, J.; King, H. F. Analyzing Molecular Static Linear Response Properties With Perturbed Localized Orbitals. *J. Chem. Phys.* **2010**, *133*, 044109–1–044109–10.
- (133) Marek, R.; Kříštková, A.; Maliňáková, K.; Toušek, J.; Marek, J.; Hocek, M.; Malkina, O. L.; Malkin, V. G. Interpretation of Indirect Nuclear Spin-Spin Couplings in Isomers of Adenine: Novel Approach to Analyze Coupling Electron Deformation Density Using Localized Molecular Orbitals. *J. Phys. Chem. A* **2010**, *114*, 6689–6700.
- (134) Thirumoorthi, R.; Chivers, T.; Vargas-Baca, I. Experimental and Theoretical Investigations of Tellurium(IV) Methanediides and Their Insertion Products With Sulfur and Iodine. *Organometallics* **2012**, *31*, 627–636.
- (135) Boughton, J. W.; Pulay, P. Comparison of the Boys and Pipek-Mezey Localizations in the Local Correlation Approach and Automatic Virtual Basis Selection. *J. Comput. Chem.* **1993**, *14*, 736–740.