

Advancing non-atom-centered basis methods for more accurate interaction energies: benchmarks and large-scale applications

Balázs D. Lőrincz^{†,‡,¶} and Péter R. Nagy^{*,†,‡,¶}

[†]*Department of Physical Chemistry and Materials Science, Faculty of Chemical Technology and Biotechnology, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

[‡]*HUN-REN-BME Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

[¶]*MTA-BME Lendület Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

E-mail: nagy.peter@vbk.bme.hu

Abstract

Recent advances in local electron correlation approaches enable the relatively routine access to CCSD(T) [that is coupled cluster (CC) with single, double, and perturbative triple excitations] computations for molecules of a hundred or more atoms. Here, approaching their complete basis set (CBS) limit becomes more challenging due to extensive basis set superposition errors, often necessitating the use of large atomic orbital (AO) basis sets with diffuse functions. Here, we study a potential remedy in the form of non-atom-centered or floating orbitals (FOs). FOs are still rarely employed even for small molecules due to the practical complication of defining their position, number, exponents, etc. The most frequently used FO method thus simply places a

single FO center with a large number of FOs toward the middle of non-covalent dimers, however, a single FO center for larger complexes can soon become insufficient. A recent alternative uses a grid of FO centers around the monomers with a single s function per center, which is currently only applicable for H, C, N, and O atoms.

Here, we build on the above advantages and mitigated some drawbacks of previous FO approaches by using a layer of FO centers and 4–9 FOs/center for each monomer. Thus, a double layer of FOs are placed between the interacting subsystems. When extending the double- ζ AO basis with this double layer of FOs, the quality of conventional augmented double- ζ or conventional triple- ζ AO bases can be reached or surpassed with less orbitals, leading to few tenths of a kcal/mol basis set errors for medium-sized dimers. This good performance extends to larger molecules (showed here up to 72 atoms), as efficient local natural orbital (LNO) CCSD(T) computations with only double- ζ AO and 4 FOs/center FO bases matches our LNO-CCSD(T)/CBS reference within ca. 0.1 kcal/mol. These developments introduce FO methods to the accurate modeling of large molecular complexes without limitations to atom types by further accelerating efficient correlation calculations, like LNO-CCSD(T).

1 Introduction

Non-covalent interactions play a major role across chemical sciences, such as in catalytic, surface, supramolecular, or biochemistry. For example, they can govern the mechanism, stereochemistry, or yield of chemical reactions, e.g., via affecting the structure and stability of transition state complexes. However, while the non-covalent interaction contributions are orders of magnitude smaller than covalent bond energies,^{1–7} their cumulative effect can be substantial, especially in extended systems.^{8–13} Hence, their accurate modeling is challenging task necessitating advanced quantum chemistry tools. For example, the wave function based electron correlation treatment could be reliable if combined with high-quality atomic orbital (AO) basis sets to approach their complete basis set (CBS) limit. Especially, the coupled

cluster (CC) model^{14–16} with single and double excitations (CCSD)¹⁷ as well as contributions from triple excitations^{18–20} can provide systematically improvable results. Here, we employ the CCSD model with perturbative triple excitations [CCSD(T)],²⁰ which is often referred to as gold standard in quantum chemistry.

While the CCSD(T) model was repeatedly shown to deliver chemical accuracy (i.e., < 1 kcal/mol errors),^{8,10,15} its steep scaling limits its applicability range conventionally to ca. 20–25 atoms even with efficient parallel implementations.^{21–28} However, recent advances relying on, e.g., natural orbital (NO) based approximations^{29–31} and their combination with local correlation approaches^{32–43} extended the limits of reliable CCSD(T) computations to 40–50^{44,45} and even to 100s of atoms.^{43,46–48} In this work, we will employ our local natural orbital (LNO) method,^{43,48–53} which enabled so far some of the most advanced CCSD(T) computations for complex intermolecular interactions. Namely, with our LNO-CCSD(T) method we reported tightly converged augmented quintuple- ζ computations for complicated supramolecular complexes of up to 132 atoms,⁵⁴ surface binding on ionic crystals matching the quality and uncertainty of experiments,⁵⁵ CBS limit interaction energies for ion-ligand complexes,⁵⁶ and quadruple- ζ level protein-ligand interaction energies up to 1023 protein atoms.^{43,48}

Hence, especially for such larger molecules, the slow convergence of the correlation energy towards the CBS limit also poses challenges. Considering interactions, one can note that the traditional AO basis sets are usually optimized for atoms, and thus can be expected to be less effective for the intermolecular region. Hence, routinely, at least triple- or quadruple- ζ AO basis sets are required, often augmented with diffuse functions for well-converged interaction energies. This can often lead to oversaturation on the atomic positions and undersaturation in the interacting region. Although some basis families offer an extensive hierarchy of systematically improving sets, increasing the number of AOs this way can rapidly lead to linear-dependency issues for large molecules. While explicit electron correlation methods can accelerate the basis set convergence also for CC computations,^{57–63} they are

most helpful to model the electron-electron cusp and may still require extensive diffuse basis sets for accurate non-covalent interactions.⁶⁴

In this study, we systematically benchmark and propose novel methods specifically designed to accelerate the basis set convergence of non-covalent interactions by using orbitals residing not only on the atomic positions. Depending on their specific purpose, such (mostly Gaussian orbital based) approaches were referred to in the literature as non-atom-centered, midbond, off-center, or floating orbital (FO) basis methods. For example, Tao and co-workers added a single midbond orbital center to the middle of the non-covalent bonds between helium and other noble gas containing non-covalent dimers.^{65–69} Extending this midbond concept, Mester and Kállay added ellipsoidal Gaussian type orbitals to the center of covalent bonds.⁷⁰ In this way, placing a midbond function halfway between two covalently bonded atoms can play the role of polarization functions, while placing midbond functions to the space between two non-covalently interacting monomers contribute to the description of intermolecular interactions. To highlight their property of not being centered on the atomic positions or the middle of bonds in all cases, we will give preference to the floating orbital denomination.

Despite their advantages, such FO basis functions are still rarely used in practice, as they have numerous additional parameters to be defined compared to the case of AO basis sets. Namely, their position, number, exponent, and angular momenta has to be determined, which in general requires a difficult, non-linear optimization procedure. The complicated task of treating these as variational parameters was taken on so far only by Császár and Tasi for a few prototypical atomic and molecular systems of 1-3 atoms.⁷¹ On top of that, the non-atom-centered basis parameters also probably depend on the quality of the AO basis set and the type of the intermolecular interactions.

Because of these complications, the relatively simple, single midbond orbital approach is applied in almost all FO-based studies.^{65–69,72–75} Due to the use of only one FO center, a relatively large FO basis is employed, that often contains at least 3 sets of s and p, 2 sets of d,

and 1 set of f basis functions (briefly 3s3p2d1f or 3321). Using this setup, Tao and co-workers carried out MP4 calculations and concluded that a smaller AO and FO basis set (than the pure AO basis set) is sufficient to reach the same accuracy in the interaction energies. This simple FO center definition was also adapted by Szalewicz and co-workers to study different, biologically relevant non-covalent complexes (S22 test set).⁷³ They calculated interaction energies in a composite MP2 and CCSD(T) scheme, and added the FO basis to reduce the AO basis set needed for CCSD(T) calculations. Patkowski and co-workers combined this single FO center approach with explicitly correlated wave function methods.^{74,76–78}

Recently, Høyvik and co-workers investigated various non-covalent interaction types, such as H-bonds, dispersion, and mixed interactions for smaller (up to 6 atoms in the A24 set) and medium sized (up to 36 atoms in the S66 set) complexes.⁷⁵ They pointed out the need for high angular momentum FOs when using a single midbond center. Compared to the above studies, Høyvik and co-workers also studied the effect of adding a second FO center for the medium-sized complexes, which resulted in slight improvement of the interaction energy accuracy. Along the line of including more FO centers, Neogrády and co-workers employed an FO center grid surrounding the entire surface of the monomers.^{79,80} Their approach places a single s type basis function on each grid point, whose parameters are optimized so far only for H, C, N, and O atoms.

In this study, we first systematically compare the performance of the single midbond and the FO center grid type methods for various non-covalent interactions of medium sized complexes up to 36 atoms, including H-bonds, ionic H-bonds, dispersion and mixed interactions.^{81,82} We also present a novel FO approach building on some of the advantageous and overcoming some of the unfavorable properties of previous FO approaches. Namely, we use more than one FO center strategically placed in a double layer formation between the monomers to cover the region of non-covalent interaction. Moreover, we found optimal compromises between the single s type and a large 3s3p2d1f bases by using a 1s1p or 1s1p1d FO basis on each FO center of the double layer. In general, this double layer FO basis can

improve cc-pVDZ interaction energies to the quality of aug-cc-pVDZ or cc-pVTZ with ca. 1.5–2 times less basis functions. Going toward larger systems of high practical relevance, the proposed double layer method is more generally applicable as it does not have atom type limitation and it overcomes the problem of diminishing FO contributions occurring with a single FO center. Moreover, the number of FOs only increases with the size of the interacting surface, which is much more favorable than the scaling of, e.g., adding diffuse AOs to all atoms. We also demonstrate the applicability of the double layer FO method in combination with our LNO-CCSD(T) method to reach larger systems. We show that the CBS limit LNO-CCSD(T)-level interaction energy of the parallelly displaced coronene dimer (72 atoms) can be approached to within ca. 0.1 kcal/mol by adding the proposed FO basis to double- or triple- ζ AO basis sets.

The paper is written as follows: in Sect. 2, we summarize each of the previously applied FO approaches in detail (Sect. 2.1-2.2), as well as introduce our novel FO method (Sect. 2.3-2.5). In Sect. 3, we introduce the technical details of the computations. In Sect. 4, we present an in-depth analysis for a few complicated systems (e.g., uracil dimer, Sect. 4.1), benchmark statistics for medium sized dimers (Sect. 4.2), and a large-scale application (Sect. 4.3).

2 Methodology

The floating orbital basis sets have the following parameters to determine: number and spatial coordinates of FO centers, angular momenta, and exponents. These parameters of the FOs could also depend on the underlying AO basis. While the AO basis optimization is already a complicated, non-linear process, it still has less degrees of freedom. The reason is that common Gaussian basis sets are mostly atom-centered and developed for each element independently (although one can note some exceptions aiming at smaller AO basis sets that were found promising for DFT interactions^{83–85}). In contrast, the number and center

position of the FOs are also unknown parameter. Additionally, due to their role in modeling non-covalent interactions, the independent, elementwise optimization of the FO basis set parameters does not seem to be an ideal strategy. All in all, the global optimization of all FO parameters, including their position, AO basis-, and molecular interaction dependence is a very challenging task, which probably contributed to the limited use of FO methods in computational chemistry. However, as shown in previous studies, it is not necessary to address all of the above complexities at once to define useful FO methods.^{65,79}

First, we discuss and analyze the properties of the existing FO methods in more detail. To that end, in Table 1 a brief summary is given on the existing FO methods^{65,79} compared to our novel approach, with more details collected in Sect. 2.1 - 2.5. In the most often employed approach of Tao and co-workers (detailed in Sect. 2.1), a single FO center is placed between the two interacting monomers (Fig. 1a).⁶⁵ In the early versions, the midpoint of the monomer center of masses was selected as FO center. Later, Szalewicz and co-workers improved the definition of this single FO center position, which we review in Sect. 2.1 and will refer to here as weighted geometric center (WGC). The most commonly applied basis placed on that center consists of 3 sets of s, 3 sets of p, 2 sets of d, and 1-1 sets of f and g functions, which will be referred to as 3s3p2d1f1g or shortly 33211, collecting 38 basis functions altogether.⁸⁶

The WGC and related single FO center methods were employed so far for relatively small monomers,⁶⁵⁻⁶⁹ most recently going up to 18 atoms.⁷⁵ This is partly explained by the previous computational limitations of CCSD(T), and thus, efforts were not yet devoted to test or extend WGC-like approaches to larger molecules. However, we anticipate that FOs on a single FO center could only cover relatively small monomers (ca. 10–20 atoms) sufficiently well, while the efficacy of a single FO center could decrease with increasing monomer size. Moreover, the small monomers appearing in WGC-like applications also had a relatively simple structure and shape compared to the complexity emerging with increasing system size. While a small monomer having a flat or relatively spherical shape has a surface simple enough for satisfactory WGC definition, the proper placement of a single (or small number

of) FO center(s) is more challenging for trimers, tetramers, etc., and for more complicated dimers, such as host-guest complexes. For example, if the host surrounds the guest molecule, the WGC definition could place the FO center somewhere within the space of the guest molecule, while the complicated shape of the surface where the host and guest interact could prevent the identification of key position(s) for the placement a single (few) FO center(s).

An alternative, second approach for the FO center definition was introduced by Neogrady and co-workers,⁷⁹ overcoming the use of only a single FO center. They placed a grid of FO centers on the surface of the interacting monomers. While they did not introduce a name for their approach, in this comparative study, we will refer to it as monomer surface grid (msG) method (detailed in Sect. 2.2). This FO center definition based on monomer surfaces is expected to be more general than WGC, as it is applicable to complexes with a wide range of size, shape, and monomer number. As they employed a relatively dense grid (with somewhat smaller grid edge length than covalent bond lengths), the number of FO centers in the msG method is approximately 3-7 times larger than the number of atoms in the dimer (Fig. 1b). The FO basis of the msG method was defined using a single s type function (1s) per msG FO center. Using this setup, they optimized the msG model parameters determining the FO positions and exponents for representative molecular dimers containing H, C, N and O atoms.⁷⁹ Therefore, currently the msG parameters are defined only for these four elements, and in the corresponding limited chemical space.

Table 1: Summary of floating orbital basis method parameters from literature studies as well as for method introduced in this work.

Floating orbital (FO) basis method	weighted geometric center (WGC) ^{65,87}	monomer surface grid (msG) ⁷⁹	double layer (DL)
FO basis	3s3p2d1f1g	1s	1s1p(1d)
Number of FO centers	1	~3-7x system size	~ interacting surface size
Number of FOs/center	38	1	4 (9)
Position of FOs	WGC of dimer	grid around monomers	interacting region
Exponents	from ref. 65	optimized in ref. 79	adapted from ref. 65
Applicability	dimers of small monomers	only H,C,N,O atoms	no atom type or size restriction

In this study, we introduce a third approach, which will be referred to as double layer (DL) method (Fig. 1e). The goal of the DL method is to extend the applicability range of previous FO methods and improve their performance. To that end, we identify the beneficial features of the WGC and msG methods, generalize and combine them with novel ideas (Table 1). In brief (details in Sect. 2.5), we place a layer of FO centers on the surface of each non-covalently interacting monomers only on the surface side facing the other monomer. The number of DL FO centers significantly extends the single center of WGC, but it is considerably smaller than the number of msG FO centers (cf. Fig. 1b and 1e), because we focus on the intermonomer region that plays an important role in the non-covalent interaction. The DL FO centers are placed so that almost all atoms on the surface have a dedicated FO center. Regarding the FO basis placed on each FO center, we appreciate the larger than 1s basis sets used in the WGC method, and the larger number and more strategically positioned FO centers of the msG approach. In the DL method, we combine these directions by using less FO centers (than in msG) with more FOs per center, including 1s1p and 1s1p1d contributions with exponents taken from the WGC method.⁶⁵ The resulting basis definition of the DL method do not require optimization and thus can be employed for more general complexes without restrictions to the elements constituting the monomers.

To better explain these aspects, next, the three investigated methods (and some variants of them) are introduced in detail in Sect. 2.1-2.5.

2.1 Weighted geometric center (WGC) method

The FO method of Tao and co-workers, which is used most often in practice defines the center of a single FO halfway between the center of mass of each subsystem.⁶⁵⁻⁶⁹ This approach was first proposed for the study of noble gas containing dimers and then was extended to monomers of up to 3-4 atoms. Turning to more complicated cases with monomers of markedly different size, such as the helium-cyanoacetylene dimer, Szalewicz and co-workers found that the FO center definition of Tao et al. could lead to a midpoint placed closer

to one of the monomers.⁸⁷ When this occurs, the FOs on the single FO center are more beneficial to the description of that closer (larger) monomer. To overcome this asymmetry, Szalewicz and co-workers proposed an improved FO center placement, where the position [$\mathbf{r}_{\text{WGC}}$ in Eq. (1)] is the r^{-6} weighted average of the midpoints of intermonomer atom pairs:

$$\mathbf{r}_{\text{WGC}} = \frac{\sum_{a \in A} \sum_{b \in B} w_{ab} \frac{\mathbf{r}_a + \mathbf{r}_b}{2}}{\sum_{a \in A} \sum_{b \in B} w_{ab}}; w_{ab} = |\mathbf{r}_a - \mathbf{r}_b|^{-6}. \quad (1)$$

Here, the w_{ab} weight is the inverse sixth power of the atom-atom distances of monomers A and B, while $\mathbf{r}_a$ and $\mathbf{r}_b$ are the spatial coordinates of atoms from subsystem A and subsystem B, respectively. To reference this definition, we will call this method here as weighted geometric center (WGC) approach.

The motivation behind the r^{-6} weights in Eq. (1) is the analogous decay of dispersion interactions between two atoms on different monomers. Consequently, larger weights are assigned to more strongly interacting atom pairs, that are closer to each other. Therefore, the definition of Eq. (1) does not favor the larger monomer and incorporates information about the monomer surfaces and their atoms. For example, in Fig. 1a, the WGC of the uracil dimer is located in the position with the smallest intermolecular distances, slightly shifted from the midpoint between the centers of the six-membered rings.

Originally, Tao and co-workers employed 3s, 3p, 2d, and 1f (shortly 3s3p2d1f or 3321) basis functions on the single FO center. The exponents of these FOs were determined in ref. 65. This 3321 FO basis was later extended with a g function by Christiansen et al. when investigating the benzene-argon dimer (resulting in 3s3p2d1f1g or shortly 33211 FO basis).⁸⁶ Due to the excellent performance of the 33211 FO basis, recently it was adapted by Høyvik and co-workers.⁷⁵ They found that for smaller systems (below 6-atom monomers) the results obtained via the single FO center is more convincing than for larger systems (below 18-atom monomers) in combination with double- ζ AO basis sets. Thus, Høyvik and co-workers also

added a second FO center manually for the larger monomers, based on chemical intuition, which lead to moderate improvements over the results with a single FO center. These limited set of results for medium-sized dimers suggest that the use of a single (or two) FO center(s) could not be sufficient when studying even larger molecular complexes.

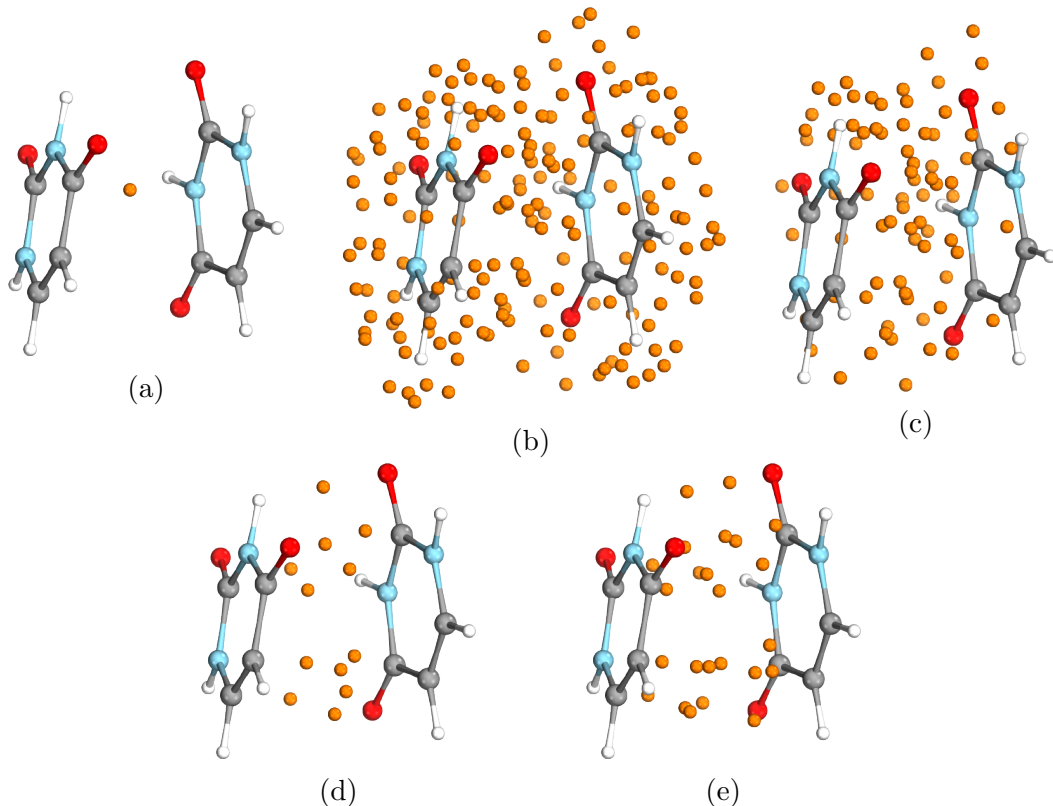


Figure 1: Position of floating orbitals (represented by the orange spheres) for the uracil dimer (24 atoms) of the S66 test set with various FO methods: weighted geometric center (a), monomer surface grid (b), interacting region grid (c), single layer (d), and double layer (e). Number of FO centers for these five methods are 1, 172, 78, 12, and 24, respectively.

2.2 Monomer surface grid (msG) of floating orbital centers

The approach of Neogrady and co-workers defines a grid of FO centers placed onto the surface of each monomer (monomer surface grid, msG method).^{79,80} The FO centers are determined in three steps for each subsystem summarized briefly as:

1. the subsystem surface is defined as the union of spheres around each atom of the subsystem with a radius of r_{msG} .

2. a uniform grid of points (with grid edge length e_{msG}) is projected onto the subsystem surface separately from the three directions defined by the principal axes of the subsystem
3. revision/removal of grid points that are too close to each other to avoid linear dependency in the combined AO and FO basis set:
 - (a) removal of one of the intramonomer grid points from those that are too close to each other
 - (b) offset of intermonomer grid point pairs that are close to each other

For example, the msG FO center list generated this way for the uracil dimer is presented in Fig. 1b.

The parameters of the msG approach (r_{msG} , e_{msG} , and the exponent of the s type FO) were optimized in ref. 79 for the H, C, N, and O atom types. The msG parameters were optimized for the interaction energies of representative systems from the S22 test set: stacked and hydrogen-bonded uracil dimer, stacked benzene-indole complex with both equilibrium and partly dissociated geometries. To simplify the optimization procedure, only two parameter sets were determined, one set for hydrogen and another set for non-hydrogen (C, N and O) atoms. However, as the FO parameters are available only for these four atom types, the current applicability of msG method is limited to the corresponding chemical space. While the FO center positions are not as simple to obtain as for the WGC approach, both the WGC and msG FO center coordinates translate and rotate with the molecular orientation, providing independence from the choice of the coordinate system.

As the optimum of e_{msG} grid edge length turned out to be smaller than 1 Å for the four atom types, usually multiple FO centers are assigned to each atom on the monomer surface. Therefore, the number of centers are 3-7 times larger than the number of atoms in the dimer. The combination of this msG FO basis with AO bases not including diffuse AOs provided similar performance for the studied interactions as the same AO basis without FOs, but

augmented with diffuse functions both in terms of numerical performance and basis set size (that is, e.g., cc-pVDZ+msG vs aug-cc-pVDZ).^{79,80}

2.3 Interacting region: space between the monomers

Considering the msG FO centers, e.g., in Fig. 1b, not every FO center is expected to be equally important for the interaction energy: those FO centers that are placed between the interacting monomers could be more important than the others, outside of the space between the monomers. To investigate this assumption further, let us identify those atoms of the surface of each subsystem that could be more important for the interaction. This interacting surface atom list will be used in Sect. 2.4 to place FO centers located only between the monomers.

As the majority of the non-covalent interaction components is expected to originate from the atoms and electrons residing on the surface of the interacting monomers, we focus on the space between the monomers, simply referred to as interacting region here. We determine the atom list of the interacting surface as follows (Fig. 2):

1. measure the minimum distance (MD) between two monomers
2. MD is multiplied by a scaling factor ($SF > 1$). This scaled minimum distance controls the spatial extent of the interacting region.
3. intermonomer atom pairs with a distance lower than $SF \cdot MD$ are selected to constitute the interacting surface.

For example, for a smaller (S) and a larger (L) fragment let s_i and l_j label the i^{th} and j^{th} atom of S and L, respectively. Then, with d_{ij} being the distance between the s_i - l_j atom pair, this atom pair is added to the interacting surface if $d_{ij} < SF \cdot MD$.

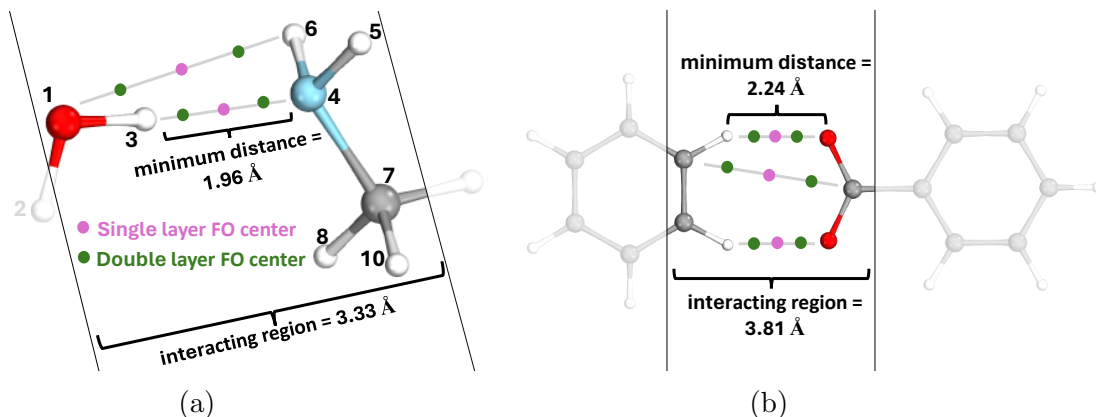


Figure 2: Illustration of the interacting surface atom list definition, as well as the positions of floating orbital centers for the single layer (pink dots) and the double layer (green dots) methods for the water–methyl-amine H-bonded complex (a), and for the benzene–benzoate ionic H-bonded complex (b).

This approach is illustrated in Fig. 2 on the examples of the water–methyl-amine hydrogen bonded complex (Fig. 2a) as well as on the benzene–benzoate ionic hydrogen bonded complex (Fig. 2b). For the sake of brevity, we only give a detailed description of the interacting surface definition for the smaller, water–methyl-amine dimer. Here, $MD = 1.96 \text{ \AA}$, which is the distance between hydrogen atom (number 3) of water and nitrogen atom (number 4) of methyl-amine. Then, the spatial extent of the interacting region, $SF \cdot MD = 1.70 \cdot 1.96 \text{ \AA}$ becomes $3.33 \text{ \AA}$, where $SF = 1.70$ was determined based on the inspection of the retained interacting surface atom list for the S66 with multiple SF choices. To the first water atom (oxygen, number 1) atoms 4, 5, and 6 of methyl-amine are closer than $SF \cdot MD$, thus first, atoms 1, 4, 5, and 6 are added to the interacting surface. Then, for atom 3 of water, atoms 7, 8, and 10 of methyl-amine are added, while none of the methyl-amine atoms are closer to atom 2 of water than $SF \cdot MD$.

2.4 Interacting region grid (irG) of floating orbital centers

Utilizing the interacting surface atom list of Sect. 2.3, we propose to restrict the FO center list of msG to an interacting region between the subsystems. To that end, we identify the msG centers that are sufficiently close to the atoms of the interacting surface in the subsystems as follows:

1. We select the subsystem with the smaller number of atoms on the interacting surface. For the i^{th} atom on the interacting surface subset of this subsystem, we find its closest intersubsystem atom pair, with distance d_{ij} . Then, we use these average of the d_{ij} distances to define the average distance between the interacting surface parts on the two subsystems (D).
2. if a msG FO center is closer to any interacting surface atom than D , then we add this FO center to the interacting region grid (shortly irG).

For example, for the uracil dimer, $D = 3.24$ Å, the number of FO centers in the irG is 78 (Fig. 1c), which is less than the number of msG FO centers in Fig. 1b by 94 centers or 55%. While the irG method considerably reduces the number of FO centers to about 3-times the number of dimer atoms, the irG centers are still closely packed. This FO center density might be necessary if only a single s type FO is placed on each irG center. Thus, in Sect. 2.5, we also investigate if less FO centers are sufficient in combination with somewhat more FOs/center.

2.5 Single and double layer of floating orbital centers

To find a more compact FO center list, let us recognize that in the irG approach, there are practically two, densely packed FO grids placed near the interacting region part on the two subsystems. Building on this, we propose two options directly placing a single and a double layer (SL and DL) of FO centers into the interacting region. The centers of the SL (DL) method are determined so that we assign roughly one (one pair of) FO center(s) to one intersubsystem atom pair as follows:

1. for the subsystem with the smaller number of atoms on the interacting surface we go through its surface atoms and find the closest intersubsystem atom pair for each. The pairing is made to be a bijection, starting with the smallest intersubsystem distance (minimum distance, MD in Fig. 2).

2. the SL (DL) FO centers are placed to the midpoints (trisection points) of the inter-subsystem atom pairs of step 1 (see Fig. 2).

SL (DL) definition is illustrated in detail in Fig. 2 on the water–methyl-amine H-bonded dimer. The first atom pair (atoms 3 and 4) corresponds to the MD. Due to the bijective construction, each atom can be utilized only once, thus atoms 1 and 6 define the second atom pair. As the smaller interacting surface (corresponding to the water molecule) consists of only 2 atoms, only two (four) SL (DL) FO centers are constructed in this example.

The choice of not using a surface atom more than once for SL/DL definition is useful to avoid close lying FO centers and to construct a relatively even distribution of FO centers. For the SL (DL) methods, there is roughly one FO center for each surface atom pair (surface atom), which is considerably less than for the irG approach. Hence, we can consider to place more than a single s type function on the SL/DL FO centers, while the total number of FOs remains similar to irG or even less than with the msG method. A potential benefit of SL over DL is the somewhat less additional FOs, although compared to the size of the AO basis set, both the SL and DL FO numbers are relatively small. On the other hand, we prefer DL over SL, especially when the dimer distance is longer (e.g., for larger monomers in Sect. 4.3 or for monomers of more irregular shape) or the monomers are, e.g., somewhat dissociated, as SL would place the FOs to a larger distance from the monomers in such cases.

In our numerical analysis, we assess FO bases of increasing size: 1s, 1s1p, 1s1p1d, 2s2p1d. We introduce the shorthand notation of method/FO basis, e.g., DL/1s1p for these combinations. For the sake of comparability with the WGC method, the exponents of these FOs were taken from ref. 65. Where there are more exponents to the same angular momentum in the 3s3p2d1f1g basis of ref. 86, the most diffuse exponents were retained. For example, when constructing the 1s1p1d FO basis building on 3s3p2d, the 0.1 exponent (in atomic unit) was retained from the s exponent list of 0.9, 0.3, and 0.1.

3 Computational details

Density fitting (DF) based conventional CCSD(T),²⁸ local MP2,^{49,50} and local natural orbital (LNO) based^{43,48,51–53} local CCSD(T) computations have been performed using the 2023 version of the MRCC quantum chemistry program suite.^{88,89} Pople type 6-31+G(2d)⁸³ and Dunning type correlation consistent basis sets^{90,91} with and without diffuse functions [(aug-)cc-pVXZ, X = D, T], as well as heavy augmented basis sets (haug-cc-pVXZ, X = D, T) were employed which are abbreviated as [(h)aug]XZ in the figures of the manuscript and the Supporting Information (SI).

DF approximation was utilized for every computation with the corresponding DF auxiliary bases for the HF⁹² and correlation energy⁹³ calculations, that is, (aug-)cc-pVXZ-RI-JK and (aug-)cc-pVXZ-RI were employed with the (aug-)cc-pVXZ AO basis set. Szalewicz and co-workers adapted the 3s3p2d1f FO basis from the work of Tao et al. and proposed the 5s5p5d4f3g DF auxiliary basis.⁹⁴ Later, Christiansen and co-workers introduced the 3s3p2f1f1g FO basis,⁸⁶ but they did not define a DF auxiliary basis corresponding to their 3s3p2d1f1g FO basis. Therefore, we extended the 5s5p5d4f3g basis with 3 set of h functions in this work, which resulted in a 5s5p5d4f3g3h auxiliary basis. Moreover, DF auxiliary bases 2s1p-RI, 3s2p1d-RI, and 4s3p2d1f-RI were constructed to fit the 1s, 1s1p, and 1s1p1d (2s2p1d) FO bases, respectively. The same FO DF auxiliary basis was employed for both the HF and the correlation energy calculations. The exponents of these FO auxiliary bases are also available in the SI.

For the complete basis set (CBS) extrapolation of the HF energies, the two-point formula suggested by Karton and Martin is used with the recommended parameters.^{95,96} Conventional and LNO based correlation energies were extrapolated with the formula of Helgaker and co-workers⁹⁷ with exponent of 2.46 (2.51) for the (aug-)cc-pV(D,T)Z extrapolation and exponent of 3 for the (aug-)cc-pV(T,Q)Z extrapolation.

The structures of the S66 compilation were taken from the original work of Řezáč, Riley, and Hobza,⁸¹ while the 21 selected ionic H-bonded dimer structures were taken from

the IHB100 test set compiled by Řezáč.⁸² The calculations were performed on equilibrium dimer structures for both the S66 and the IHB100 test set. The names of the selected 21 ionic H-bonded dimers are listed in the SI. We utilized MP2-F12/aug-cc-pV(T,Q)Z-F12 and MP2/aug-cc-pV(Q,5)Z interaction energies as CBS references for the S66 and IHB100 DF-MP2 interaction energies, respectively.^{82,98}

To characterize the performance of the different FO methods, the mean absolute error (MAE), the root mean square deviation (RMSD), and the maximum absolute error (MAX) were calculated. The timing measurements are performed with a single 64-core AMD EPYC 7763 processor.

4 Results and discussion

The accuracy assessment and comparison of the previous and here introduced FO methods depend on a large variety of aspects. Besides the noted FO parameters (number and position of FO centers, and basis set parameters, like exponents), such a benchmark work should consider the AO basis combined with the FO method types, the level of electron correlation treatment, as well as representative molecule and non-covalent interaction types. Here, we consider the following:

1. FO methods: weighted geometric center (WGC), monomer surface grid (msG), interacting region grid (irG), single layer (SL) and double layer (DL)
2. FO basis sets: 1s, 1s1p, 1s1p1d, 2s2p1d, 3s3p2d1f1g
3. AO basis sets: 6-31+G(2d), (h)aug-cc-pVXZ and cc-pVXZ, X = D, T
4. wave function methods: MP2, CCSD(T)
5. non-covalent interaction types: hydrogen bond, dispersion, mixed, ionic hydrogen bond

The number of variables to be considered is too large to explore and discuss here all possible combinations. Therefore, first we show in detail three of the most complicated dimers of the

S66 test set,⁸¹ that is the uracil dimer ($\pi - \pi$ stacking), the uracil base pair, as well as the benzene-peptide dimer to illustrate the more important trends, and to decrease the number of setting combinations to be assessed (Sect. 4.1). Then, a comprehensive statistical analysis is presented for the more practical setting combinations on the S66 test set as well as on 21 ionic hydrogen bonded complexes selected from the IHB100 test set⁸² (Sect. 4.2). Finally, a large-scale practical application is presented in Sect. 4.3.

4.1 Numerical performance of FO methods: uracil dimer

We start the analysis with the dispersion dominated ($\pi - \pi$ stacking) uracil dimer from the S66 test set, as this system exhibits the largest, more than 5 kcal/mol basis set incompleteness error (BSIE) in the interaction energy with the cc-pVDZ basis set. Moreover, we carried out similar studies on the uracil base pair, and the benzene-peptide dimer, which results are available in the SI (Fig. S1 and S2). Additionally, we found that the basis set errors are analogous for MP2 and CCSD(T). Here, the CCSD(T) results are presented (e.g., in Fig. 3 for the $\pi - \pi$ stacking uracil dimer), while the analogous MP2 results are available in the SI (Fig. S2). We also combined the FO methods with the 6-31+G(2d) AO basis set,⁸³⁻⁸⁵ which results are available in Table S1 in the SI.

The improvement of the BSIE for various FO methods with respect to the CBS limit reference [that is CP corrected CCSD(T)/aug'-cc-pV(T,Q)Z] is demonstrated in Fig. 3, in combination with the cc-pVDZ (left panel), and the cc-pVTZ AO basis sets (right panel). Here, we show on the x axis the total number of AO and FO basis functions relative to the size of cc-pVDZ. Apparently, the rate of improvement due to the increasing number of additional FO functions is similar for cc-pVDZ, cc-pVTZ, and 6-31+G(2d). Therefore, we will mainly focus on the results with the cc-pVDZ AO basis.

Starting with the most conventional WGC/3s3p2d1f1g FO method, the error of interaction energy decreases by 40% for the dispersion dominated uracil dimer (green square in Fig. 3) and decreases by 50 and 60% for the uracil base pair and benzene-peptide dimer,

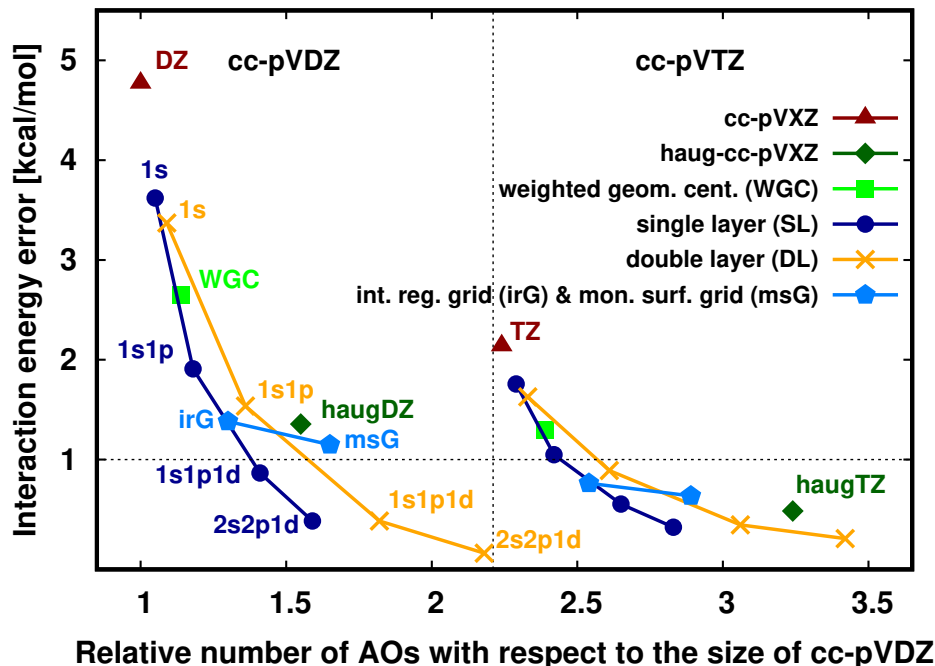


Figure 3: DF-CCSD(T)/cc-pVXZ (+FO) (X = D, T) correlation energy contribution error of CP corrected interaction energies as a function of the combined AO and FO basis set size for the π - π stacked uracil dimer of the S66. Reference: CP corrected DF-CCSD(T)/haug-cc-pV(T,Q)Z.⁹⁸

respectively (green squares in Fig. S1a and S1b) compared to the pure cc-pVDZ results. The monomer surface grid method already achieves haug-cc-pVDZ accuracy for all the three investigated dimers (c.f. msG vs haugDZ in Fig. 3 and S1) in line with the results of ref. 79. Moreover, the cc-pVDZ + msG results outperform those with the cc-pVTZ AO basis set, using 1.3–1.4 times less basis functions for all three cases, but the improvement gained via the msG method is somewhat more pronounced for the π - π stacked uracil dimer and for the benzene-peptide dimer, than for the uracil base pair. Compared to the msG results, retaining only the FOs in the interacting region, the irG approach leads to negligible loss of accuracy with about half of the basis functions, for all three systems.

Turning to the analysis of the single layer (SL) and the double layer (DL) FO methods (dark blue and orange curves in Fig. 3, S1, and S2 respectively), the effect of systematically adding higher angular momentum functions in the interacting region can also be studied. The convergence trend towards the CBS limit is analogous for all three dimers, when comparing

the SL and the DL FO methods, therefore, we will mainly discuss the DL results. Starting with the π - π stacked uracil dimer, its interaction energy error with the pure AO basis decreases by only 20% when DL/1s is employed. By adding DL/1s1p to cc-pVDZ, the accuracy of cc-pVTZ, while with DL/1s1p1d even the accuracy of haug-cc-pVTZ is reached with only 61% and 56% of the basis functions, respectively. The interaction energy error decreases even further with DL/2s2p1d method, however, the majority of the improvement occurs already with 1s1p and 1s1p1d. This implies that the further extension of the FO basis might not lead to notable improvements, at least in combination with double- and triple- ζ AO basis sets. Finally, comparing the FO methods to each other, DL/1s1p and irG considerably outperform WGC by providing haug-cc-pVDZ quality with less orbitals than in haug-cc-pVDZ. The msG moderately, while DL/1s1p1d significantly improves that, the latter outperforms even cc-pVTZ + WGC, and matching haug-cc-pVTZ and cc-pVTZ + msG with significantly smaller number of basis functions altogether.

Inspecting the results of the uracil base pair, the improvement gained via the SL and DL method is less significant with respect to the pure cc-pVDZ AO basis set than for the π - π stacked uracil dimer. For example, to reach the accuracy of the pure cc-pVTZ or haug-cc-pVDZ AO basis set, at least the cc-pVDZ + DL/1s1p1d FO basis is necessary, while none of the here introduced FO basis added to cc-pVDZ seems to approach the quality of the pure haug-cc-pVTZ. The dependence of the SL and DL results on the interaction types can be attributed to the different sizes of the interacting surface atom lists. Namely, the near parallel placement of the monomers in the π - π stacking uracil dimer results half of the dimer atoms (12 atoms in this case) in the interacting surface atom list. Compared to that, for the uracil base pair, only the two H-bond donor and two acceptor atoms are added to the interacting surface atom list, which brings less FOs to the interacting region altogether, thus improving BSIEs to a smaller extent.

4.2 Statistical analysis on non-covalent dimer test sets

We continue with a statistical analysis on the representative S66 set, separated into its conventional H-bonding, dispersion, and mixed (polar-, dispersion, etc.) interaction subsets. We found it important to extend the benchmark set with stronger interactions and shorter intermolecular distances to make sure that FO centers are not too close to each other in such cases. To that end, we selected 21 representative ionic H-bonded complexes from the IHB100 test set, referred to as IHB100/21.⁸² Due to the similarity of the basis set convergence of the MP2 and CCSD(T) interaction energies for the uracil dimer (c.f. Fig. 3 and S2), we expect that an MP2-based statistical analysis on S66 and IHB100/21 will also be representative for the case of CCSD(T). Therefore, to enable the assessment of a large number of AO and FO combinations, we continue with an MP2-based error analysis.

First, we investigate the performance of the DL method in Fig. 4 with respect to increasing FO basis size. Here, we present the distribution of interaction energy errors for the DL FO method added to the cc-pVDZ (Fig. 4a), cc-pVTZ (Fig. 4b), and aug-cc-pVDZ (Fig. S3c of the SI) conventional AO basis sets, respectively. (The analogous single layer results are collected in Fig. S3a-S3b of the SI.) Similarly to the case of the uracil dimer (Sect. 4.1), the convergence towards the CBS limit when adding more FOs is analogous with the cc-pVDZ, cc-pVTZ, and also with the aug-cc-pVDZ AO basis sets for each interaction type, therefore, we will discuss in detail only the case of cc-pVDZ (Fig. 4a).

For the hydrogen bonded dimers of S66 (bottom panels of Fig. 4), two peaks are found on the error distribution curves, around 2 and 4 kcal/mol. The second peak at 4 kcal/mol can be attributed to overcorrections caused by the Counterpoise correction for double H-bonded systems of S66: uracil base pair, acetic acid dimer, acetamide dimer, acetic acid-uracil dimer, as well as acetamide-uracil dimer. This overcorrection has been pointed out for H-bonded systems of S66 previously.⁹⁸ The size of the BSIE with cc-pVDZ is decreased by the additional FOs, diffuse functions, and the increased AO basis set (c.f. aug-cc-pVDZ in Fig. 4a, and cc-pVTZ or aug-cc-pVTZ in Fig. 4b), however, this second peak remains also with

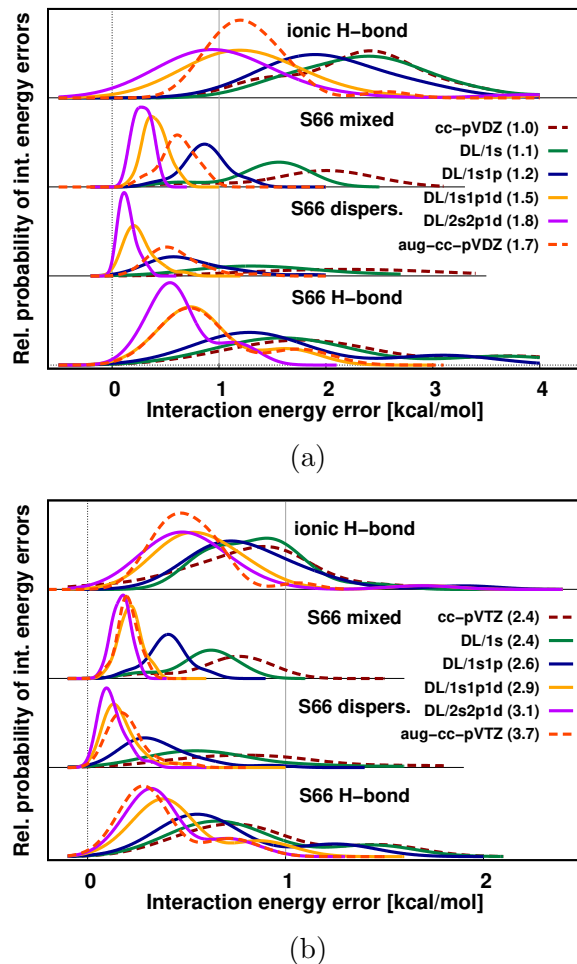


Figure 4: Relative probability of interaction energy errors with the double layer FO method and 1s, 1s1p, 1s1p1d, and 2s2p1d FO bases separated to four subsets. (The analogous single layer FO results are plotted in Fig. S3 of the SI.) Level of theory: DF-MP2/(aug-)cc-pVXZ, with $X = D$ in a, and $X = T$ in b. The total number of basis functions relative to the size of cc-pVDZ is collected in parentheses besides the basis set labels.

the largest basis set employed here (aug-cc-pVTZ).

Extending the pure cc-pVXZ AO basis (dark red dashed curves in Fig. 4) with diffuse functions (light red dashed curves in Fig. 4) as well as with FO basis of increasing size (solid curves in Fig. 4) affects non-covalent bond types differently: the H-bonded dimers in S66 and IHB100/21 form one, and the dispersion dominated and mixed systems of S66 form a second set. Comparing the AO basis and the DL/1s performance (dark red dashed and green curves of Fig. 4), we again find a marginal improvement, which is somewhat more pronounced for the dispersion/mixed subsets. Adding the DL/1s1p FOs (blue curves) to

cc-pVDZ (cc-pVTZ) already brings the peaks of the error distributions to around 1 (around 0.5) kcal/mol, with somewhat larger errors remaining for the H-bonded cc-pVDZ + DL/1s1p cases. These numerical results can also be studied in Table 2 and in Table S2 of the SI, where mean absolute error (MAE) values are collected for the four molecular subsets, as well as for the various AO and FO settings separately. A second step of significant improvement comes from adding the d functions of DL/1s1p1d (orange curves). Here, one finds 0.91 (1.41) and 0.48 (0.63) kcal/mol MAEs with cc-pVDZ + DL/1s1p1d and cc-pVTZ + DL/1s1p1d settings, respectively for H-bonds (ionic H-bonds) in the top of Table 2 (bottom of Table S2 of the SI). Similarly significant improvement is observed for dispersion (mixed) interactions, where 0.26 (0.41), as well as 0.17 (0.31) kcal/mol MAEs characterize the numerical performances of the cc-pVDZ + DL/1s1p1d and cc-pVTZ + DL/1s1p1d, respectively, as seen in the bottom of Table 2 (top of Table S2 of the SI). The cc-pVDZ + FO results outperform the pure cc-pVTZ errors for most dimers already with DL/1s1p, and for the ionic H-bonds with DL/1s1p1d. Moreover, the cc-pVDZ + DL/1s1p1d basis delivers comparable performance to aug-cc-pVDZ for the H-bonded subsets and outperforms aug-cc-pVDZ AO basis for the dispersion/mixed subsets with a somewhat smaller basis set in average. Compared to that, again, the improvement brought by DL/2s2p1d (purple curves in Fig. 4) is systematic, but relatively small.

Inspecting the number of FOs added to the dimers of H-bonded subsets, and to the other two S66 subsets provides an additional layer of understanding to their somewhat different behavior. Namely, the interacting surface for the H-bonded dimers often consists of the H-bond donor and acceptor atoms (often resulting in 2 atoms for single and 4 atoms for double H-bonds interacting surfaces). Thus, the corresponding FO basis of 1s and 1s1p contain not much more than 2-8 FOs. Compared to that, the 5 FOs per center extension brought by the d set of 1s1p1d is significant. In contrast, the dispersion/mixed interacting surfaces are larger, which yields more FO centers. Moreover, the dispersion interactions are significantly weaker, which often means smaller absolute BSIE. The contribution of the above factors

provides additional explanation to why we observe faster convergence towards the CBS limit reference with increasing FO size for the dispersive/mixed subsets than for the H-bonded dimers.

As the DL method (especially with the 1s1p and the 1s1p1d basis) is the more accurate protocol compared to the SL in Fig. S4 of the SI, the numerical performance of DL/1s1p and DL/1s1p1d will be investigated in Fig. 5 more thoroughly, while we set aside the 1s, 2s2p1d, and larger FO basis set options. In Fig. 5, we compare the performance of different AO basis sets (cc-pVXZ and aug-cc-pVXZ; $X = D, T$ in Fig. 5a) with that of the FO methods extending the cc-pVDZ (Fig. 5c), aug-cc-pVDZ (Fig. 5b), and cc-pVTZ (Fig. 5d) AO basis sets for the same four non-covalent interaction subsets. Let us briefly consider first only the pure AO basis sets. For the H-bonded systems (top and bottom panels of Fig. 5a), the interaction energies improve more if the cardinal number is increased (from $X = D$ to $X = T$, c.f. blue and green curves), while for the dispersion/mixed subsets (two middle panels) the interaction energies improve more by adding diffuse functions (c.f. blue and red curves).

Turning to the analysis of the WGC method (dark red curve) for H-bonded subsets, it has comparable numerical performance to irG, msG, and DL/1s1p1d methods (orange, purple, and blue dashed curves, respectively) for both the cc-pVDZ (Fig. 5c) and the cc-pVTZ cases (Fig. 5d). Moreover, combining these FO methods with the cc-pVDZ AO basis set, the accuracy of cc-pVTZ can be reached. Compared to that, the aug-cc-pVDZ + WGC/33211 combination (Fig. 5b) slightly outperforms the other methods, probably due to the small size (up to 24 atoms) of the studied H-bonded dimers. In contrast to the case of the H-bonded subsets, we found a notably smaller improvement of the interaction energy errors with cc-pVDZ + WGC/33221 for dispersion/mixed systems, as those subsets contain larger monomers and larger interacting surfaces. For comparison, the switching from cc-pVDZ AO basis set to aug-cc-pVDZ and cc-pVTZ (without FOs) for these dispersion (mixed) interactions already reduces the MAEs from 2.45 (1.83) kcal/mol to 0.60 (0.61) and 0.89 (0.70) kcal/mol, respectively (see Tables 2 and S2). Thus, the MAE of 1.08 (0.64) kcal/mol

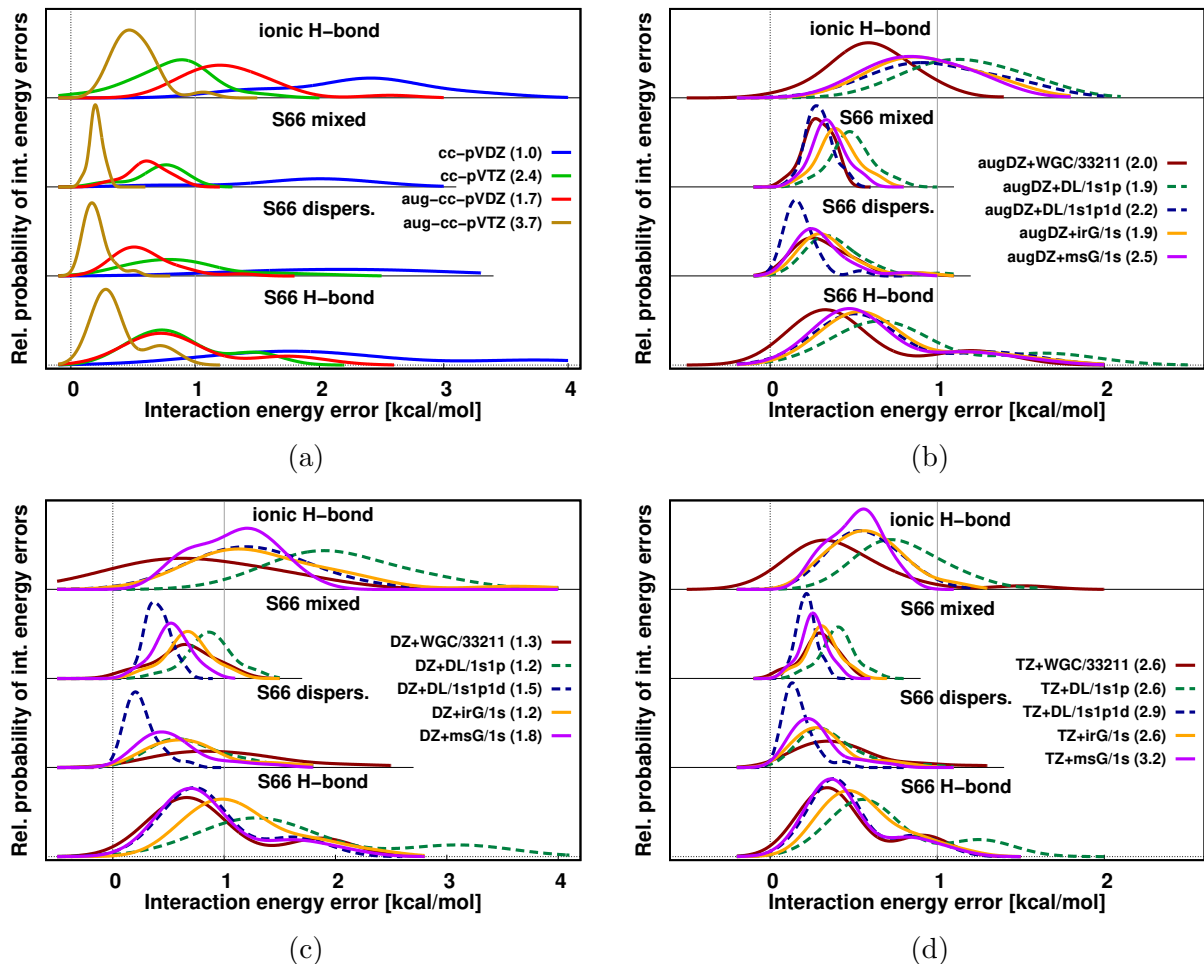


Figure 5: Relative probability of interaction energy errors at DF-MP2/(aug-)cc-pVXZ, X=D, T level of theory using pure AO basis sets (a) and various FO bases added to cc-pVDZ (DZ, c), aug-cc-pVDZ (augDZ, b), and cc-pVTZ (TZ, d). The total number of basis functions relative to the size of cc-pVDZ is collected in parentheses besides the basis set labels.

for cc-pVDZ + WGC/33211 is outperformed by pure aug-cc-pVDZ and cc-pVTZ, but adding the WGC to aug-cc-pVDZ or cc-pVTZ still almost halves their MAEs (Table 2 and S2).

Considering the performance of the irG and msG methods in Fig. 5c-5d (orange and purple curves), the maxima of their error distribution curves are already within ca. 0.2 kcal/mol of each other with cc-pVDZ AO basis. This is achieved with 1.5 times less basis functions in the irG approach than the msG for each molecular subset. This difference in their local maxima is decreasing further with larger AO basis sets. These results again suggest that the most significant improvement can be achieved via adding FO centers to the interacting

region first. We can also observe that the irG and msG performances are closer to each other for dispersion/mixed interactions, than for H-bonds, which also can be explained by the larger spatial extent of the interacting region. Since the interacting surface atom lists are more extended for dispersion/mixed systems, more FO centers are retained for these than for H-bonded dimers. Therefore, the BSIE values of the dispersive and mixed subsets with the irG method can become smaller.

Investigating the double layer method for dispersion/mixed interactions, the results with cc-pVDZ + DL/1s1p (dashed green curves in Fig. 5) are comparable to those with the irG and msG methods, since sufficient number of FO centers are added to the interacting region. This trend remains for the case of aug-cc-pVDZ and cc-pVTZ as well. Regarding the H-bonded systems with the smaller number of atoms on the interacting surfaces, DL/1s1p FOs bring smaller improvement, which trend is the most notable for cc-pVDZ and decreases with larger AO basis sets (aug-cc-pVDZ or cc-pVTZ). The next step of improvement comes from adding d type FOs. Namely, cc-pVDZ + DL/1s1p1d (dashed blue curves in Fig. 5) approaches the accuracy of aug-cc-pVTZ with 2.5 times less basis functions on the average for dispersion/mixed interactions. Moreover, it is as good as or often notably outperforms msG and irG for all studied AO basis sets. Furthermore, the DL/1s1p1d approach provides comparable results to irG and msG methods for the H-bonding subsets.

As we can observe in Fig. 5c-5d (and also in Fig. S5a-S5f of the SI), the error distribution curves for different FO methods overlap with each other more often as we extend the AO basis set size. The corresponding MAE values show the same, monotonic decrease (c.f. DZ, augDZ, TZ, and augTZ columns of Table 2). Since the smaller aug-cc-pVTZ errors are not visible well at this scale, we present their MAE and MAX values in bar charts (Fig. 6). As these error measures are very close to each other with and without the additional FOs, here the efficiency of FO methods with larger, e.g., aug-cc-pVTZ AO basis set appears to decrease. Namely, the improvements for all four investigated non-covalent interaction types and all FO methods is below ca. 0.1 kcal/mol with respect to the pure aug-cc-pVTZ results.

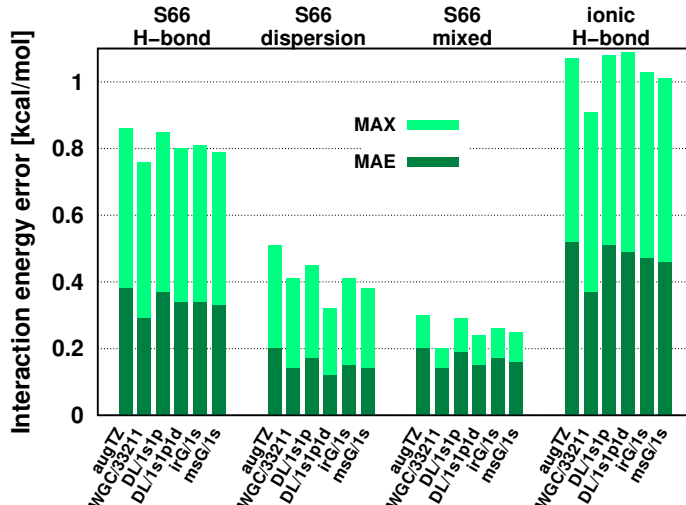


Figure 6: DF-MP2/aug-cc-pVTZ correlation energy contribution error of CP corrected interaction energies for the studied FO methods.

Here, the diffuse AOs in aug-cc-pVTZ probably span a similar space as the relatively low angular momentum orbitals in the FO basis, and thus the already small errors with respect to the CBS reference do not decrease considerably. Since the errors are already at the few tenths of a kcal/mol range, studying the effect of higher angular momentum FOs is set aside for future work.

We look more closely at the MAEs of the H-bonded and dispersion dimers of the S66 in Table 2, while the analogous mixed and ionic H-bonded subset results are available in Table S2 of the SI. The MAEs also show that adding FOs to AO basis sets without diffuse AOs provides diffuse basis set quality results, especially with the msG, irG, and DL/1s1p1d FO methods. For example, cc-pVDZ + FO methods match pure aug-cc-pVDZ and cc-pVTZ results (c.f. columns DZ–TZ of Table 2). Similarly, cc-pVTZ + FO results are comparable to the quality of aug-cc-pVTZ. These observations hold for WGC/33211 only for the H-bonded and for DL/1s1p only for the dispersion/mixed cases.

Table 2: Mean absolute basis set errors [kcal/mol] for the H-bonded (top) and dispersion dominated (bottom) subsets of S66. Level of theory: DF-MP2/(aug-)cc-pVXZ; X = D, T.

	DZ	augDZ	TZ	augTZ	(D,T)Z	aug(D,T)Z
H-bonds of S66						
AO basis	2.18	0.95	0.88	0.38	0.12	0.06
WGC/33211	0.92	0.52	0.46	0.29	0.19	0.16
SL/1s1p	1.79	0.90	0.74	0.37	0.13	0.08
SL/1s1p1d	1.35	0.75	0.59	0.35	0.15	0.13
DL/1s1p	1.68	0.87	0.71	0.37	0.14	0.08
DL/1s1p1d	0.91	0.69	0.48	0.34	0.23	0.15
irG/1s	1.18	0.70	0.56	0.34	0.18	0.14
msG/1s	0.92	0.64	0.47	0.33	0.21	0.15
dispersion interactions of S66						
AO basis	2.45	0.60	0.89	0.20	0.10	0.03
WGC/33211	1.08	0.33	0.44	0.14	0.07	0.03
SL/1s1p	1.02	0.45	0.45	0.18	0.12	0.03
SL/1s1p1d	0.52	0.28	0.25	0.14	0.10	0.06
DL/1s1p	0.73	0.40	0.36	0.17	0.14	0.04
DL/1s1p1d	0.26	0.18	0.17	0.12	0.11	0.08
irG/1s	0.71	0.36	0.33	0.15	0.10	0.04
msG/1s	0.55	0.30	0.27	0.14	0.11	0.05

We also carried out CBS extrapolations for AO basis sets with and without diffuse functions (last two columns of Table 2 and Fig. 7). Partly because dispersion/mixed interactions are weaker than H-bonds in the studied cases, the corresponding cc-pV(D,T)Z [aug-cc-pV(D,T)Z] interaction energies are better converged for dispersion/mixed subsets than for H-bonds. Namely, in Fig. 7, the MAEs for dispersion/mixed interaction energies are within 0.15 [0.10] kcal/mol, while for H-bonds, they are within 0.25 [0.20] kcal/mol, respectively. The extrapolated interaction energies improved only moderately or not at all using the presented FO methods with respect to the pure AO basis extrapolations (either in terms of MAX or MAE values).

Regarding the CBS extrapolations at the cc-pV(D,T)Z level, let us recall that the conventional inverse cubic extrapolation formula turned out to be less effective for smaller AO basis sets due to the slow correlation energy convergence towards the CBS limit at this basis set size.⁹⁷ Therefore, empirical optimization of the extrapolation exponent $\gamma = 3$ was recom-

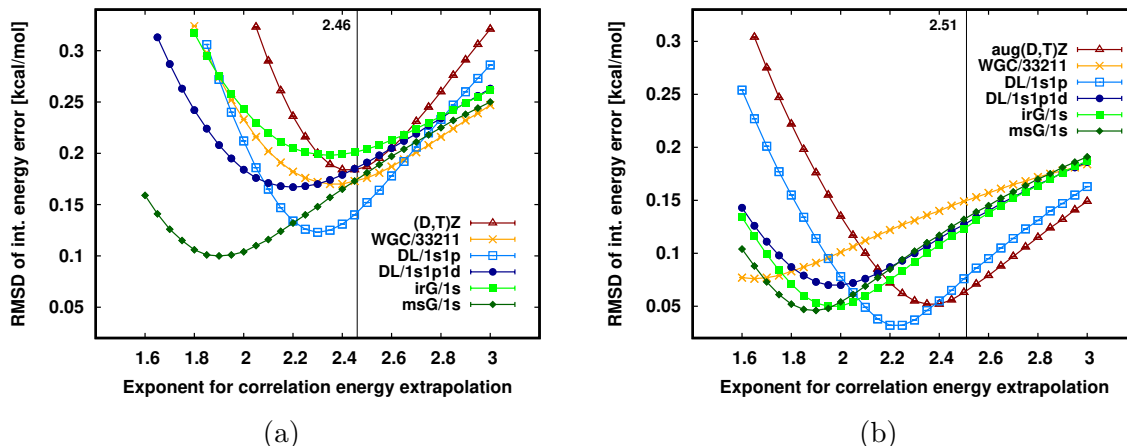


Figure 8: CBS extrapolation exponent scans for the cc-pV(D,T)Z [(D,T)Z, a] and aug-cc-pV(D,T)Z [aug(D,T)Z, b] extrapolation with various FO methods. Vertical lines at 2.46 and 2.51 on the x axis represent the previously recommended cc-pV(D,T)Z and aug-cc-pV(D,T)Z extrapolation exponents.⁹⁷

4.3 Large-scale application of the FO methods

In the previous sections, we focused on medium sized molecular dimers (up to 36 atoms) to assess previous and here proposed FO methods. However, one goal of the here introduced FO methods is the applicability to large molecules of practical interest. Therefore, we investigated the interaction energy convergence of the parallelly displaced coronene dimer (72 atoms, Fig. 9), as this system requires diffuse, at least triple- or quadruple- ζ quality atom-centered basis sets and Counterpoise corrections for well-converged interaction energies according to our comparisons to aug-cc-pV(Q,5)Z level computations.⁵⁴ However, while feasible with our LNO-CCSD(T) implementation, already the aug-cc-pVQZ computations are quite demanding for large molecules.

We generated the FO center lists of various FO methods (see, e.g., the irG and DL FO centers in Fig. 9a-9b) to investigate their effect on the basis set convergence of the coronene dimer interaction energy. To that end, LNO-CCSD(T)/cc-pVXZ ($X = D, T$) interaction energies with Tight LNO thresholds (Fig. 10 and Table 3) were compared to our aug-cc-pV(Q,5)Z CBS reference computations from ref. 54.

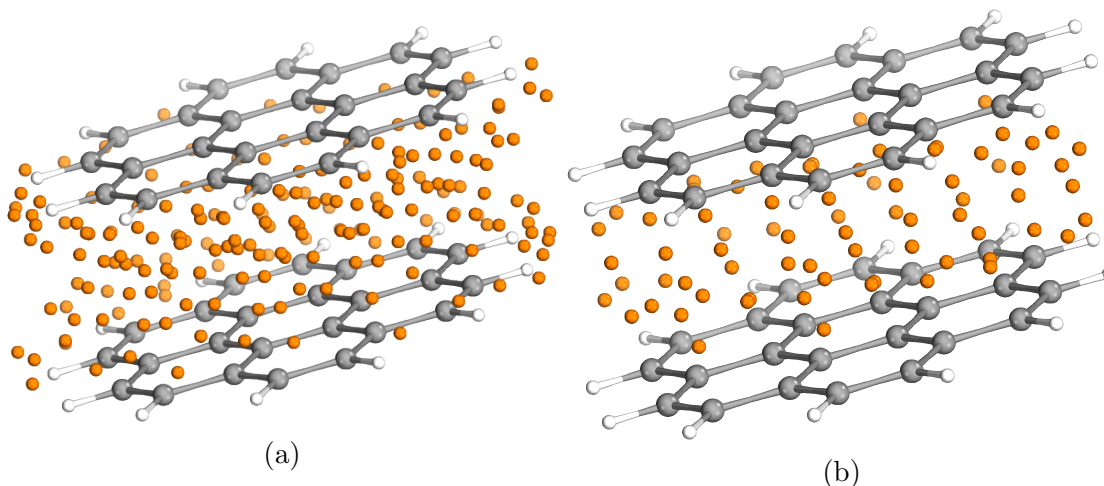


Figure 9: Position of the interacting region grid (irG, a) and the double layer (DL, b) FO centers for the parallelly displaced coronene dimer (represented by the orange spheres). Number of FO centers are 191 and 72, respectively.

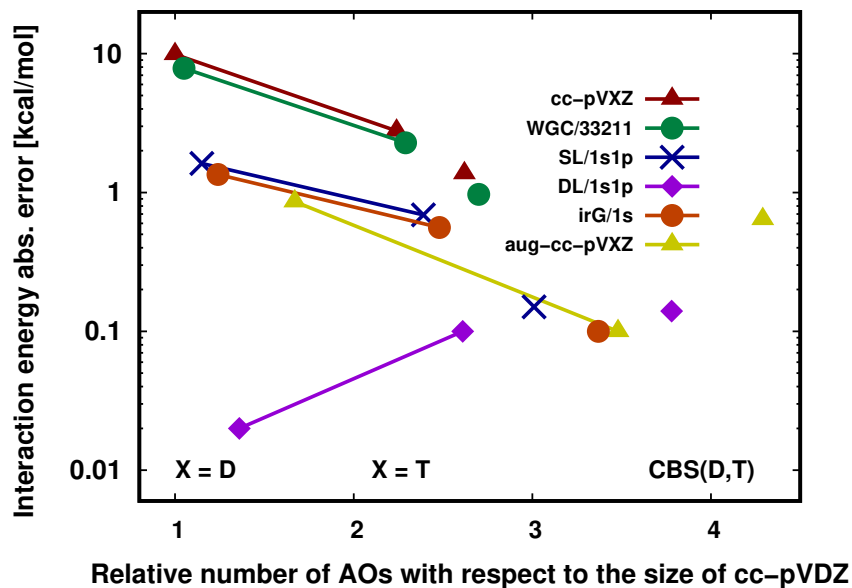


Figure 10: Absolute interaction energy errors of the LNO-CCSD(T)/(aug-)cc-pVXZ ($X = D, T$) for the CP corrected correlation energy contributions in a logarithmic scale. Test system: parallelly displaced coronene dimer. Reference: CP corrected LNO-CCSD(T)/aug-cc-pV(Q,5)Z.⁵⁴

Inspecting Fig. 10, we find the pure AO basis and WGC results similar (c.f. red and green curves). The reason is that compared to the extended size of the coronene dimer, the effect of a single FO center even with 38 extra functions is not notable. Hence, the time of the computation, as well as the number of the basis functions do not change considerably

as well, when adding the WGC FOs to either cc-pVDZ or cc-pVTZ (see the corresponding columns of Table 3 for the numerical results). With the irG method (orange in Fig. 10), the errors of the interaction energy improve by almost one order of magnitude with respect to the pure cc-pVDZ AO basis results. It requires only 1.2 times increase in the total AO and FO basis set size (relative to cc-pVDZ) and ca. twice as long computation time. This finding also holds for the SL/1s1p FO method (blue in Fig. 10) with not notably less basis functions with respect to irG/1s. Both the SL/1s1p and the irG/1s numerical performances are similar to the pure aug-cc-pVDZ basis (yellow in Fig. 10), but the number of basis functions used for the aug-cc-pVDZ is ca. 1.5 times larger than that of the SL/1s1p or irG/1s FO methods.

The performance of the cc-pVDZ + DL/1s1p FO method (purple in Fig. 10) is better than that of irG/1s and SL/1s1p, as even 0.1 kcal/mol accuracy is surpassed, which is 2 orders of magnitude improvement in the interaction energy relative to the pure cc-pVDZ error. Considering the 0.1 kcal/mol error of cc-pVTZ + DL/1s1p and cc-pV(D,T)Z CBS results, the excellent cc-pVDZ + DL/1s1p performance should be interpreted as partly fortunations. While cc-pVTZ + DL/1s1p considerably outperforms cc-pVTZ + irG/1s and cc-pVTZ + SL/1s1p, their CBS extrapolated results are all great, showing ca. 0.1 kcal/mol remaining BSIE. Moreover, cc-pVTZ + DL/1s1p is highly competitive with the pure aug-cc-pVTZ AO basis set, as both exhibit only 0.1 kcal/mol absolute BSIE, but the aug-cc-pVTZ basis set contains almost 1.4 times more basis functions. Due to the planar and parallel structure of the coronene dimer (Fig. 9), all dimer atoms were added to the interacting surface atom list. Thus, with the DL method, 72 FO centers were positioned in the interacting region and thus 1 set of s and p FOs are assigned to each atoms. We note that further increasing the FO basis to 1s1p1d resulted in near linear dependency of the basis set, which originates from the planar structure of the monomers and the highly ordered alignment of the corresponding AOs and FOs. Therefore, we did not perform further investigations with DL/1s1p1d on the coronene dimer, while it could still be a useful approach for large, but not as symmetric molecular dimers.

Turning to the analysis of the msG method (see Table 3), its numerical performance is similar to that of DL/1s1p both in terms of accuracy and computation time as it contains only 134 more FOs than DL/1s1p. Furthermore, we can also observe that the improvement is larger going from the pure cc-pVXZ basis set to cc-pVXZ + irG/1s than from cc-pVXZ + irG/1s to cc-pVXZ + msG/1s. This finding also corroborates that the FO centers outside of the interacting region bring less significant improvement to the interaction energies, than the 45% of the msG FOs kept in the interacting region when constructing irG. All in all, the performance of DL/1s1p and msG/1s with both cc-pVDZ and cc-pVTZ (as well as of irG/1s with cc-pVTZ) are remarkable, being in the tenths of a kcal/mol BSIE range. Since we cannot generalize far from these promising results obtained for one challenging system, the broader investigation of large non-covalent dimers is needed and planned in the future.

We note that Tight LNO thresholds were selected for the coronene dimer investigations to suppress the local approximation error to or below the range of BSIE. Therefore, due to the increased number of operations induced by the Tight LNO threshold, as well as the complicated long-range $\pi - \pi$ interactions in the coronene dimer, the calculations are more time consuming than the average for 50–100 atom molecules.⁴⁸ Compared to that, LNO-CCSD(T)/cc-pVDZ (cc-pVTZ) calculations with Normal LNO settings, without FOs took 13.3 (37.2) hours, while DL/1s1p with the same theoretical level and LNO threshold took 24.0 (57.6) hours on the dimer, both with only 8 CPU cores.

Table 3: LNO-CCSD(T)/(aug-)cc-pVXZ (X = D, T) interaction energy errors of the parallelly displaced coronene dimer with respect to the CP corrected LNO-CCSD(T)/aug-cc-pV(Q,5)Z CBS reference.⁵⁴ Local correlation threshold: Tight. The cc-pVDZ (cc-pVTZ) calculations were executed on 24 (32) CPU cores, the aug-cc-pVDZ (aug-cc-pVTZ) calculation was executed on 32 (40) CPU cores.

AO basis	FO method (num. of centers)	(Rel.) num. of AOs + FOs	Int. energy error [kcal/mol]	Runtime (dimer) [h]
cc-pVDZ	– (0)	792 (1.0)	9.89	45.7
	WGC/33211 (1)	830 (1.1)	7.84	54.4
	SL/1s1p (36)	908 (1.2)	1.62	81.9
	irG/1s (191)	983 (1.2)	1.35	90.9
	DL/1s1p (72)	1080 (1.4)	0.02	126.4
	msG/1s (422)	1214 (1.5)	0.08	125.6
cc-pVTZ	– (0)	1776 (2.2)	2.78	145.4
	WGC/33211 (1)	1814 (2.3)	2.28	165.0
	SL/1s1p (36)	1892 (2.4)	0.69	188.8
	irG/1s (191)	1967 (2.5)	0.56	172.0
	DL/1s1p (72)	2064 (2.6)	0.10	209.4
	msG/1s (422)	2198 (2.8)	0.05	209.0
cc-pV(D,T)Z	– (0)	–	-1.37	–
	WGC/33211 (1)	–	-0.97	–
	SL/1s1p (36)	–	0.15	–
	irG/1s (191)	–	0.10	–
	DL/1s1p (72)	–	0.14	–
	msG/1s (422)	–	0.03	–
aug-cc-pVDZ	– (0)	1320 (1.7)	0.86	82.6
aug-cc-pVTZ	– (0)	2760 (3.5)	-0.10	206.1
aug-cc-pV(D,T)Z	– (0)	–	-0.64	–

5 Conclusions and outlook

In this study, we systematically compared previous non-atom-centered or floating orbital (FO) basis approaches^{65,79} with here proposed novel FO methods for medium-sized (H-bond, ionic H-bond, dispersion, and mixed) molecular complexes,^{81,82} as well as on a large-scale application (coronene dimer). The so far almost exclusively employed approach uses a single FO center in the space between the two interacting monomers combined with a relatively large basis (up to 38 FOs) placed on that center.⁶⁵ While even a single FO center can decrease the double- ζ AO basis set errors by 50-60%, e.g., for dispersion dominated dimers of ca. 20–

30 atoms, the analogous improvement for more extended molecules is much smaller (e.g., ca. 20% for the coronene dimer). Overcoming some of the limitations of using a single FO center, a recent alternative method places a grid of FO centers and a single s type function per grid point onto the surface of the monomers.⁷⁹

Here, we first showed that there is no need for completely surrounding the monomers with FOs, their use can be limited to the space between the non-covalently interacting monomers. This resulted in the use of up to 1.5 times less AO and FO basis functions altogether compared to the FO grid method completely surrounding the monomers with negligible loss of accuracy. Furthermore, we introduced a novel FO method which strategically adds one-one layer of FO centers onto the surface of each monomer facing the other monomer, thereby adding ca. one FO center to each atom that play a key role in the interaction. With this more compact FO center list, we could employ more than a single s function on this double layer of FO centers. Moreover, this new double layer approach is considerably more general regarding some limitations of the previous methods. For example, it performs much better for large molecules than the single FO center and does not have the limitation of the previous FO center grid method being optimized only for H, C, N, and O atoms.⁷⁹

Our statistical analysis showed that the most beneficial choice is to use a set of s and p functions (1s1p) or an additional set of d functions (1s1p1d) per FO center in the new double layer FO method. For example, for smaller, H-bonded systems, cc-pVDZ with the 1s1p1d FO basis could outperform (reach) the accuracy of the pure cc-pVTZ (aug-cc-pVDZ) AO basis with 1.6 (1.1) times less (AO and FO) basis functions. For the more dispersion dominated complexes with larger interacting surfaces, even the accuracy of the pure aug-cc-pVTZ is approached (with 2.1 times less orbitals), while cc-pVDZ with even the smaller 1s1p FO basis had pleasing performance.

Therefore, the proposed FO method can successfully replace diffuse AOs or decrease the cardinal number of the AO basis, both of which are particularly helpful for large molecules to ease the computational cost and frequent near-linear dependency issues. Additionally,

number of FOs added to the interacting surface scales much more favorably with system size than adding, e.g., diffuse AO onto all atoms. As presented for the complicated interactions in the coronene dimer, the combination of FO methods with efficient asymptotically linear-scaling local correlation methods, such as our local natural orbital LNO-CCSD(T)^{43,48,52} can make large-scale interaction energy computations accurate and routinely accessible. In particular for the coronene dimer, adding an FO extension of ca. 40% of the size of cc-pVDZ with our novel FO method could decrease the 10 (3) kcal/mol basis set error of the cc-pVDZ (cc-pVTZ) interaction energies to ca. 0.1 kcal/mol compared to the expensive LNO-CCSD(T)/aug-cc-pV(Q,5)Z CBS extrapolated reference.⁵⁴

Considering that our LNO-CCSD(T) method was applicable to compute protein-ligand interaction energies using more than a 1000 atoms and a quadruple- ζ AO basis set,^{43,48} the presented FO method development can notably extend the scope of accurate and routinely accessible non-covalent interaction computations. For example, LNO-CCSD(T) with triple- ζ AO and the proposed FO basis should be routinely applicable for few hundred atoms with a single CPU and few 10 GBs of memory, covering a wide range of supramolecular, catalyst-substrate, drug-protein active site, solute-solvent, surface adsorption, etc. interactions, which will be further demonstrated also in our forthcoming studies.

Acknowledgement

The work is supported by the National Research, Development, and Innovation Office (NK-FIH, Grant No. FK142489), the New National Excellence Program of the Ministry for Culture and Innovation (ÚNKP-23-3-I-BME-205 and ÚNKP-23-5-BME-408), the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and the ERC Starting Grant No. 101076972, “aCCuracy”. The research reported in this paper is part of project BME-EGA-02, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed un-

der the TKP2021 funding scheme. The computing time granted on the Hungarian HPC Infrastructure at NIIF Institute, Hungary, is gratefully acknowledged.

Supporting Information Available

- further numerical data visualization regarding the accuracy assessment and comparison of the here introduced floating orbital methods (PDF)
- complete list of the names and the Cartesian coordinates employed for the S66 and IHB100/21 complexes, as well as for the parallelly displaced coronene dimer augmented with floating orbital positions
- MRCC input file (MINP) samples for the DF-CCSD(T) calculation of the S66 uracil dimer, serial number 26
- MRCC basis set file (GENBAS) collecting the utilized floating orbital basis sets, contraction coefficients, and exponents
- complete list of the calculated HF interaction energies and the DF-MP2, and if available, DF-CCSD(T), LMP2, LNO-CCSD, and LNO-CCSD(T) correlation energy contributions of the interaction energies (XLS)

References

- (1) Kollman, P. A. Noncovalent interactions. *Acc. Chem. Res.* **1977**, *10*, 365.
- (2) Stone, A. J. *The Theory of Intermolecular Forces*; International Series of Monographs on Chemistry; Clarendon Press, 1997.
- (3) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
- (4) Černý, J.; Hobza, P. Non-covalent interactions in biomacromolecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5291.
- (5) Hobza, P.; Müller-Dethlefs, K. *Non-covalent interactions: Theory and Experiment*; Royal Society of Chemistry: Cambridge, 2010.
- (6) Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. Revealing Noncovalent Interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498.
- (7) Riley, K. E.; Hobza, P. Noncovalent interactions in biochemistry. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 1.
- (8) Al-Hamdani, Y. S.; Tkatchenko, A. Understanding non-covalent interactions in larger molecular complexes from first principles. *J. Chem. Phys.* **2019**, *150*, 010901.
- (9) Klimeš, J.; Michaelides, A. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **2012**, *137*, 120901.
- (10) Řezáč, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016**, *116*, 5038.
- (11) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105.

- (12) Hermann, J.; DiStasio, R. A. J.; Tkatchenko, A. First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications. *Chem. Rev.* **2017**, *117*, 4714.
- (13) Gray, M.; Herbert, J. Density functional theory for van der Waals complexes: Size matters. *ChemRxiv* **2024**,
- (14) Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966**, *45*, 4256.
- (15) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291.
- (16) Bartlett, R. J. Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Ann. Rev. Phys. Chem.* **1981**, *32*, 359.
- (17) Purvis III, G. D.; Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **1982**, *76*, 1910.
- (18) Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. Towards a full CCSDT model for electron correlation. *J. Chem. Phys.* **1985**, *83*, 4041.
- (19) Bartlett, R. J.; Sekino, H.; Purvis III, G. D. Comparison of MBPT and coupled-cluster methods with full CI. Importance of triplet excitation and infinite summations. *Chem. Phys. Lett.* **1983**, *98*, 66.
- (20) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479.
- (21) Deumens, E.; Lotrich, V. F.; Perera, A.; Ponton, M. J.; Sanders, B. A.; Bartlett, R. J.

- Software design of ACES III with the super instruction architecture. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 895.
- (22) Anisimov, V. M.; Bauer, G. H.; Chadalavada, K.; Olson, R. M.; Glenski, J. W.; Kramer, W. T. C.; Aprà, E.; Kowalski, K. Optimization of the Coupled Cluster Implementation in NWChem on Petascale Parallel Architectures. *J. Chem. Theory Comput.* **2014**, *10*, 4307.
- (23) Pitoňák, M.; Aquilante, F.; Hobza, P.; Neogrády, P.; Noga, J.; Urban, M. Parallelized implementation of the CCSD(T) method in MOLCAS using optimized virtual orbitals space and Cholesky decomposed two-electron integrals. *Collect. Czech. Chem. Commun.* **2011**, *76*, 713.
- (24) Janowski, T.; Pulay, P. Efficient Parallel Implementation of the CCSD External Exchange Operator and the Perturbative Triples (T) Energy Calculation. *J. Chem. Theory Comput.* **2008**, *4*, 1585.
- (25) Peng, C.; Calvin, J. A.; Valeev, E. F. Coupled-cluster singles, doubles and perturbative triples with density fitting approximation for massively parallel heterogeneous platforms. *Int. J. Quantum Chem.* **2019**, *119*, e25894.
- (26) Shen, T.; Zhu, Z.; Zhang, I. Y.; Scheffler, M. Massive-parallel Implementation of the Resolution-of-Identity Coupled-cluster Approaches in the Numeric Atom-centered Orbital Framework for Molecular Systems. *J. Chem. Theory Comput.* **2019**, *15*, 4721.
- (27) Hummel, F.; Tsatsoulis, T.; Grüneis, A. Low rank factorization of the Coulomb integrals for periodic coupled cluster theory. *J. Chem. Phys.* **2017**, *146*, 124105.
- (28) Gyevi-Nagy, L.; Kállay, M.; Nagy, P. R. Integral-direct and parallel implementation of the CCSD(T) method: Algorithmic developments and large-scale applications. *J. Chem. Theory Comput.* **2020**, *16*, 366.

- (29) 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.
- (30) Taube, A. G.; Bartlett, R. J. Frozen natural orbitals: Systematic basis set truncation for coupled-cluster theory. *Collect. Czech. Chem. Commun.* **2005**, *70*, 837.
- (31) Nagy, P. R.; Gyevi-Nagy, L.; Kállay, M. Basis set truncation corrections for improved frozen natural orbital CCSD(T) energies. *Mol. Phys.* **2021**, *119*, e1963495.
- (32) Jin, Y.; Bartlett, R. J. Perturbation Improved Natural Linear-Scaled Coupled-Cluster Method and Its Application to Conformational Analysis. *J. Phys. Chem. A* **2018**, *123*, 371.
- (33) Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151.
- (34) Scuseria, G. E.; Ayala, P. Y. Linear scaling coupled cluster and perturbation theories in the atomic orbital basis. *J. Chem. Phys.* **1999**, *111*, 8330.
- (35) Li, W.; Piecuch, P. Multilevel Extension of the Cluster-in-Molecule Local Correlation Methodology: Merging Coupled-Cluster and Møller–Plesset Perturbation Theories. *J. Phys. Chem. A* **2010**, *114*, 6721.
- (36) Kobayashi, M.; Nakai, H. Divide-and-conquer-based linear-scaling approach for traditional and renormalized coupled cluster methods with single, double, and noniterative triple excitations. *J. Chem. Phys.* **2009**, *131*, 114108.
- (37) Li, W.; Ni, Z.; Li, S. Cluster-in-molecule local correlation method for post-Hartree–Fock calculations of large systems. *Mol. Phys.* **2016**, *114*, 1447.
- (38) Friedrich, J.; Walczak, K. Incremental CCSD(T)(F12)|MP2-F12—A Method to Obtain

- Highly Accurate CCSD(T) Energies for Large Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 408.
- (39) Masur, O.; Schütz, M.; Maschio, L.; Usvyat, D. Fragment-based direct-local-ring-coupled-cluster doubles treatment embedded in the periodic Hartree–Fock solution. *J. Chem. Theory Comput.* **2016**, *12*, 5145.
- (40) Feldt, M.; Mata, R. A. Hybrid Local Molecular Orbital: Molecular Orbital Calculations for Open Shell Systems. *J. Chem. Theory Comput.* **2018**, *14*, 5192.
- (41) Schütz, M.; Yang, J.; Chan, G. K.-L.; Manby, F. R.; Werner, H.-J. The orbital-specific virtual local triples correction: OSV-L(T). *J. Chem. Phys.* **2013**, *138*, 054109.
- (42) Usvyat, D.; Maschio, L.; Schütz, M. Periodic and fragment models based on the local correlation approach. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1357.
- (43) Nagy, P. R.; Kállay, M. Approaching the basis set limit of CCSD(T) energies for large molecules with local natural orbital coupled-cluster methods. *J. Chem. Theory Comput.* **2019**, *15*, 5275.
- (44) Gyevi-Nagy, L.; Kállay, M.; Nagy, P. R. Accurate reduced-cost CCSD(T) energies: Parallel implementation, benchmarks, and large-scale applications. *J. Chem. Theory Comput.* **2021**, *17*, 860.
- (45) Kállay, M.; Horváth, R. A.; Gyevi-Nagy, L.; Nagy, P. R. Basis set limit CCSD(T) energies for extended molecules via a reduced-cost explicitly correlated approach. *J. Chem. Theory Comput.* **2023**, *19*, 174.
- (46) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]. *J. Chem. Phys.* **2018**, *148*, 011101.

- (47) Ma, Q.; Werner, H.-J. Explicitly correlated local coupled-cluster methods using pair natural orbitals. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1371.
- (48) Nagy, P. R. State-of-the-art local correlation methods enable accurate and affordable gold standard quantum chemistry up to a few hundred atoms. *Chem. Sci.* **2024**, *15*, 14556.
- (49) Nagy, P. R.; Samu, G.; Kállay, M. An integral-direct linear-scaling second-order Møller–Plesset approach. *J. Chem. Theory Comput.* **2016**, *12*, 4897.
- (50) Szabó, P. B.; Csóka, J.; Kállay, M.; Nagy, P. R. Linear scaling open-shell MP2 approach: algorithm, benchmarks, and large-scale applications. *J. Chem. Theory Comput.* **2021**, *17*, 2886.
- (51) Nagy, P. R.; Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: Redundancy-free triples correction using Laplace transform. *J. Chem. Phys.* **2017**, *146*, 214106.
- (52) Nagy, P. R.; Samu, G.; Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: Improved algorithm and benchmark applications. *J. Chem. Theory Comput.* **2018**, *14*, 4193.
- (53) Szabó, P. B.; Csóka, J.; Kállay, M.; Nagy, P. R. Linear-scaling local natural orbital CCSD(T) approach for open-shell systems: algorithm, benchmarks, and large-scale applications. *J. Chem. Theory Comput.* **2023**, *19*, 8166.
- (54) Al-Hamdani, Y. S.; Nagy, P. R.; Barton, D.; Kállay, M.; Brandenburg, J. G.; Tkatchenko, A. Interactions between large molecules pose a puzzle for reference quantum mechanical methods. *Nat. Commun.* **2021**, *12*, 3927.
- (55) Shi, B.; Zen, A.; Kapil, V.; Nagy, P. R.; Grüneis, A.; Michaelides, A. A coming of age

- for many-body methods: Achieving consensus with experiments for CO on MgO. *J. Am. Chem. Soc.* **2023**, *145*, 25372.
- (56) Delgado, J. M.; Nagy, P. R.; Varma, S. Polarizable AMOEBA Model for Simulating Mg^{2+} -Protein-Nucleotide Complexes. *J. Chem. Inf. Model.* **2024**, *64*, 378.
- (57) Klopper, W.; Manby, F. R.; S.Ten-no; Valeev, E. F. R12 methods in explicitly correlated molecular electronic structure theory. *Int. Rev. Phys. Chem.* **2006**, *25*, 427.
- (58) Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. Explicitly Correlated Electrons in Molecules. *Chem. Rev.* **2012**, *112*, 4.
- (59) Werner, H.-J.; Adler, T. B.; Knizia, G.; Manby, F. R. In *Recent Progress in Coupled Cluster Methods*; Čársky, P., Paldus, J., Pittner, J., Eds.; Springer, 2010; p 573.
- (60) Kong, L.; Bischoff, F. A.; Valeev, E. F. Explicitly correlated R12/F12 methods for electronic structure. *Chem. Rev.* **2012**, *112*, 75.
- (61) Grüneis, A.; Hirata, S.; Ohnishi, Y.-y.; Ten-no, S. Perspective: Explicitly correlated electronic structure theory for complex systems. *J. Chem. Phys.* **2017**, *146*, 080901.
- (62) Tew, D. P.; Klopper, W.; Neiss, C.; Hättig, C. Quintuple- ζ quality coupled-cluster correlation energies with triple- ζ basis sets. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1921.
- (63) Kállay, M.; Horváth, R. A.; Gyevi-Nagy, L.; Nagy, P. R. Size-consistent explicitly correlated triple excitation correction. *J. Chem. Phys.* **2021**, *155*, 034107.
- (64) Sylvetsky, N.; Kesharwani, M. K.; Martin, J. M. L. The aug-cc-pVnZ-F12 basis set family: Correlation consistent basis sets for explicitly correlated benchmark calculations on anions and noncovalent complexes. *J. Chem. Phys.* **2017**, *147*, 134106.
- (65) Tao, F.-M.; Pan, Y.-K. Møller–Plesset perturbation investigation of the He_2 potential and the role of midbond basis functions. *J. Chem. Phys.* **1992**, *97*, 4989.

- (66) Tao, F.-M.; Pan, Y.-K. An accurate ab initio calculation of Ne₂ potential. *Chem. Phys. Lett.* **1992**, *194*, 162.
- (67) Tao, F.-M.; Pan, Y.-K. Ab initio potential energy curves and binding energies of Ar₂ and Mg₂. *Mol. Phys.* **1993**, *81*, 507.
- (68) Tao, F.-M. The use of midbond functions for ab initio calculations of the asymmetric potentials of HeNe and HeAr. *J. Chem. Phys.* **1993**, *98*, 3049.
- (69) Tao, F.-M.; Klemperer, W. Accurate ab initio potential energy surfaces of ArHF, ArH₂O, and ArNH₃. *J. Chem. Phys.* **1994**, *101*, 1129.
- (70) Mester, D.; Csontos, J.; Kállay, M. Unconventional bond functions for quantum chemical calculations. *Theor. Chem. Acc.* **2015**, *134*, 74.
- (71) Tasi, G.; Császár, A. G. Hartree–Fock-limit energies and structures with a few dozen distributed Gaussians. *Chem. Phys. Lett.* **2007**, *438*, 139.
- (72) Ding, Y.; Mei, Y.; Zhang, J. Z. H.; Tao, F.-M. Efficient Bond Function Basis Set for $\pi - \pi$ Interaction Energies. *J. Comput. Chem.* **2008**, *29*, 275.
- (73) Podeszwa, R.; Patkowski, K.; Szalewicz, K. Improved interaction energy benchmarks for dimers of biological relevance. *Phys. Chem. Chem. Phys.* **2010**, *12*, 5974.
- (74) Dutta, N. N.; Patkowski, K. Improving “Silver Standard” Benchmark Interaction Energies with Bond Functions. *J. Chem. Theory Comput.* **2018**, *14*, 14.
- (75) Matveeva, R.; Falck Erichsen, M.; Koch, H.; Høyvik, I.-M. The effect of midbond functions on interaction energies computed using MP2 and CCSD(T). *J. Comput. Chem.* **2022**, *43*, 121.
- (76) Patkowski, K. Basis set converged weak interaction energies from conventional and explicitly correlated coupled-cluster approach. *J. Chem. Phys.* **2013**, *138*, 154101.

- (77) Kodrycka, M.; Patkowski, K. Platinum, gold, and silver standards of intermolecular interaction energy calculations. *J. Chem. Phys.* **2019**, *151*, 070901.
- (78) Kodrycka, M.; Patkowski, K. Efficient Density-Fitted Explicitly Correlated Dispersion and Exchange Dispersion Energies. *J. Chem. Theory Comput.* **2021**, *17*, 1435.
- (79) Melicherčík, M.; Pitoňák, M.; Kellö, V.; Hobza, P.; Neogrády, P. Off-Center Gaussian Functions, an Alternative Atomic Orbital Basis Set for Accurate Noncovalent Interaction Calculations of Large Systems. *J. Chem. Theory Comput.* **2013**, *9*, 5296.
- (80) Melicherčík, M.; Suchá, D.; Neogrády, P.; Pitoňák, M. Off-center Gaussian functions: Applications toward larger basis sets, post-second-order correlation treatment, and truncated virtual orbital space in investigations of noncovalent interactions. *Int. J. Quantum Chem.* **2018**, *118*, e25580.
- (81) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427.
- (82) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.* **2020**, *16*, 2355.
- (83) Chan, B. Optimal Small Basis Set and Geometric Counterpoise Correction for DFT Computations. *J. Chem. Theory Comput.* **2023**, *19*, 3958.
- (84) Chan, B.; Dawson, W.; Nakajima, T. Sorting drug conformers in enzyme active sites: the XTB way. *Phys. Chem. Chem. Phys.* **2024**, *26*, 12610.
- (85) Chan, B.; Ho, J. Counterpoise correction from a practical perspective: is the result worth the cost? *Aust. J. Chem.* **2023**, *76*, 864.
- (86) Koch, H.; Fernández, B.; Christiansen, O. The benzeneargon complex: A ground and excited state ab initio study. *J. Chem. Phys.* **1998**, *108*, 2784.

- (87) Akin-Ojo, O.; Bukowski, R.; Szalewicz, K. Ab initio studies of He-HCCCN interaction. *J. Chem. Phys.* **2003**, *119*, 8379.
- (88) Kállay, M.; Nagy, P. R.; Mester, D.; Rolik, Z.; Samu, G.; Csontos, J.; Csóka, J.; Szabó, P. B.; Gyevi-Nagy, L.; Hégyel, B. et al. The MRCC program system: Accurate quantum chemistry from water to proteins. *J. Chem. Phys.* **2020**, *152*, 074107.
- (89) Kállay, M.; Nagy, P. R.; Mester, D.; Gyevi-Nagy, L.; Csóka, J.; Szabó, P. B.; Rolik, Z.; Samu, G.; Csontos, J.; Hégyel, B. et al. MRCC, *a quantum chemical program suite*. See <https://www.mrcc.hu/> **Accessed Aug 1, 2024**,
- (90) Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007.
- (91) Kendall, R. A.; Dunning Jr., T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796.
- (92) Weigend, F. Hartree–Fock Exchange Fitting Basis Sets for H to Rn. *J. Comput. Chem.* **2008**, *29*, 167.
- (93) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175.
- (94) Podeszwa, R.; Bukowski, R.; Szalewicz, K. Potential Energy Surface for the Benzene Dimer and Perturbational Analysis of $\pi - \pi$ Interactions. *J. Phys. Chem. A* **2006**, *110*, 10345.
- (95) Karton, A.; Martin, J. M. L. Comment on: “Estimating the Hartree–Fock limit from finite basis set calculations”. *Theor. Chem. Acc.* **2006**, *115*, 330.
- (96) Neese, F.; Valeev, E. F. Revisiting the Atomic Natural Orbital Approach for Basis Sets:

Robust Systematic Basis Sets for Explicitly Correlated and Conventional Correlated ab initio Methods? *J. Chem. Theory Comput.* **2011**, *7*, 33.

- (97) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106*, 9639.
- (98) Nagy, P. R.; Gyevi-Nagy, L.; Lőrincz, B. D.; Kállay, M. Pursuing the basis set limit of CCSD(T) non-covalent interaction energies for medium-sized complexes: case study on the S66 compilation. *Mol. Phys.* **2023**, *121*, e2109526.