Representation of the virtual space in extended systems – a correlation energy convergence study

A. S. Hansen, G. Baardsen, E. Rebolini, L. Maschio & T. B. Pedersen

Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS | Check for updates |

Representation of the virtual space in extended systems – a correlation energy convergence study

A. S. Hansen [a], G. Baardsen[a], E. Rebolini [a,b], L. Maschio [c] and T. B. Pedersen [a]

[a]Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Oslo, Oslo, Norway; [b]Institut Laue Langevin, Grenoble, France; [c]Dipartimento di Chimica, Universitá di Torino, Torino, Italy

ABSTRACT
We present an investigation of the convergence behaviour of the local second-order Møller-Plesset perturbation theory (MP2) correlation energy toward the canonical result for three insulating crystals with either projected atomic orbitals (PAOs) or various orthonormal representations of the virtual orbital space. Echoing recent results for finite molecular systems, we find that significantly fewer PAOs than localised orthonormal virtual orbitals are required to reproduce the canonical correlation energy. We find no clear-cut correlation between conventional measures of orbital locality and the ability of the representation to span the excitation space of local domains. We show that the PAOs of the reference unit cell span parts of the excitation space that can only be reached with distant local orthonormal virtual orbitals.
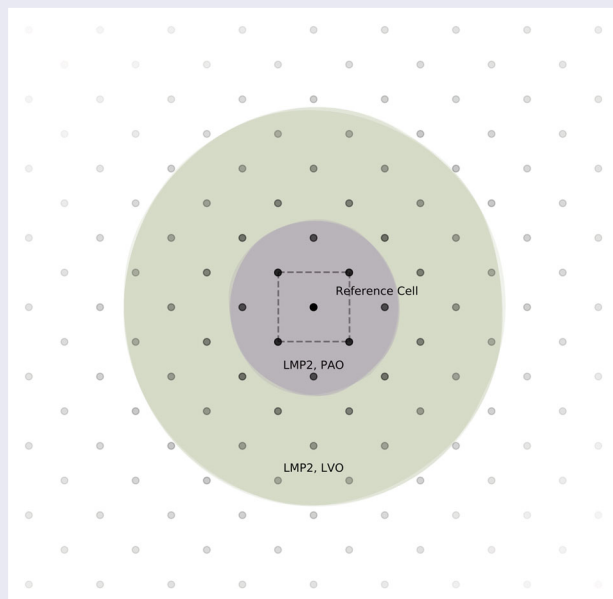
1. Introduction

Orbital localisation is a powerful tool in molecular and solid-state quantum chemistry, providing both intuitively appealing interpretations and visualisations of electronic structure [1–3], and opportunities for efficient implementations of correlated theories. Spatially localised orbitals can be obtained either by a suitably chosen unitary transformation of the inherently delocalised canonical Hartree-Fock (HF) or Kohn-Sham orbitals [4] or directly through a restrained noncanonical self-consistent optimisation [5]. From an algorithmic perspective, localised orbitals pave the way for the implementation of orbital-based embedding/fragment schemes [6, 7] and for exploiting the short-range character of electron correlation effects to greatly reduce the

CONTACT A. S. Hansen ✉ a.s.hansen@kjemi.uio.no 🖳 Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Oslo, P.O. Box 1033 Blindern, Oslo N-0315, Norway

computational complexity of post-HF methods [8–10]. In this work, we will be concerned with the latter aspect.

The concept of 'localized orbitals' is not uniquely defined, however, and different localisation functionals have been proposed. Among the oldest and most widely used proposals are the Foster-Boys [11, 12], Pipek-Mezey [13], and Edmiston-Ruedenberg [14, 15] functionals, which produce localised orbitals that minimise the orbital spread, maximise the orbital partial Mulliken charges over as few atoms as possible, and maximise the Coulomb self-repulsion of the orbitals, respectively. Neither of these locality measures is directly related to the electron-correlation problem, i.e. they are not defined through a quantity that directly enters an expression for the correlation energy. Consequently, it is far from obvious which localisation functional leads to the most efficient general-purpose local-correlation algorithm.

Local electron-correlation methods rely on at least two features:

(1) The occupied and virtual orbitals should be confined within a small volume in space [16], such that the differential overlap between them decay as rapidly as possible with distance.
(2) At the same time, the virtual orbitals should be constructed in such a way that a very small number of them is sufficient to accurately represent the excitation space of each pair of occupied orbitals.

The latter has been tackled mainly through the introduction of orbital-specific [17] or pair-specific [18] virtual orbitals. These sets are constructed from an initial set of localised virtual orbitals by means of first-order estimates of the correlation amplitudes – that is, they are constructed directly from features of the electron correlation effects of the system at hand.

Much effort has been devoted to point (1), albeit using measures not directly related to electron correlation effects. It is an open question whether nonorthonormal local orbitals are preferable to orthonormal ones [19–21]. Relaxing the orthogonality constraint leads to simpler nodal structures, which, in particular, may lead to smoother and more rapid decay of the tails of the functions. At the same time, however, the resulting set will be linearly dependent, which one might fear can lead to a larger number of orbitals required to span a given subspace. The most widely-used approach in local correlation theories is to choose an orthonormal set for the occupied space, obtained through localisation of the occupied HF orbitals, while nonorthogonal, linearly dependent projected atomic orbitals (PAOs) are used to represent the virtual manifold. The PAOs are easily computed from the converged HF density matrix and, historically, the choice

of PAOs as virtual orbital basis has more to do with the lack of robust algorithms for the localisation of orthonormal virtual orbitals than superior performance in local correlation treatments. With the robust localisation algorithms developed in recent years by Høyvik, Jørgensen, and coworkers [4, 22–24], orthonormal virtual-orbital localisation can be performed reliably and efficiently, warranting a comparison of the performance of PAOs versus localised orthonormal virtual orbitals (LVOs) in local correlation treatments.

Recently, Werner and coworkers [25–27] reported that PAOs outperform LVOs in molecular calculations. While the LVOs are more local than PAOs in terms of orbital spread, significantly fewer PAOs than LVOs are required in the excitation domains to recover the same fraction of the exact correlation energy. The authors did not present a definitive reason for this somewhat counterintuitive result, however.

The central role of locality in electron correlation treatments is even more pertinent in 3D periodic systems, where the canonical orbitals are forced by translation symmetry to be delocalised over the entire infinite solid. The dense packing of 3D periodic systems enhances the effectiveness of locality-based screening procedures, leading to even more pronounced computational savings than for finite molecular systems.

In this work, we compare the performance of PAOs and different sets of LVOs for local second-order Møller-Plesset (MP2) theory of representative insulators: the covalent diamond crystal, the ionic lithium hydride crystal, and the molecular prussic acid crystal.

2. Theoretical background

The electronic correlation energy of a weakly correlated system can be written in the coupled-cluster (CC) formalism as [28]

$$E_c = \sum_{ij} \sum_{ab} (t_{ij}^{ab} - t_i^a t_j^b)(2(ia \mid jb) - (ib \mid ja)), \quad (1)$$

where we have assumed a closed-shell system for simplicity. We use latin letters i, j, k to denote occupied spatial orbitals and a, b, c to denote virtual spatial orbitals obtained from a preceeding HF calculation. With real orbitals, the electron repulsion integrals (ERIs) are defined as

$$(ia \mid jb) = \iint \frac{\varphi_i(\boldsymbol{r})\varphi_a(\boldsymbol{r})\varphi_j(\boldsymbol{r}')\varphi_b(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, d\boldsymbol{r}' \, d\boldsymbol{r}. \quad (2)$$

The single- and double-excitation amplitudes, t_i^a and t_{ij}^{ab}, respectively, are determined from a nonlinear equation system that, for an n-electron system, may involve up

to n-tuple excitations, depending on the chosen CC model. The simplest model is second-order Møller-Plesset (MP2) theory [28] where the single-excitation amplitudes vanish due to the Brillouin condition. The double-excitation amplitudes are obtained from the set of equations [23]

$$r_{ij}^{ab} = \sum_c \left(f_{ac} t_{ij}^{cb} + t_{ij}^{ac} f_{cb} \right) - \sum_k \left(f_{ik} t_{kj}^{ab} + t_{ik}^{ab} f_{kj} \right)$$
$$+ (ia \mid jb) = 0, \tag{3}$$

where f is the Fock matrix and all orbitals are assumed to be orthonormal. Solving the MP2 equations scales as $O(N^5)$ with N a measure of system size such as the number of atom-centred basis functions used to expand the occupied and virtual orbitals. The MP2 energy, Equation (1), and amplitude equations, Equation (3), are invariant to rotations among occupied and among virtual orbitals separately. For insulators, this can be exploited to bring forth significant sparsity in the Fock matrix and ERIs which, in turn, imply sparsity in the amplitudes. This observation dates back to Pulay [8] and forms the basis for linear-scaling implementations that today approach a near-black-box level of sophistication [23, 27, 29–33].

Linear scaling can be achieved as the following simple argument shows. Given suitably localised occupied and virtual orbitals, the integral $(ia \mid jb)$ will be negligible unless a is centred in the vicinity of i in the sense that the product $\varphi_i(\mathbf{r})\varphi_a(\mathbf{r})$ must be non-vanishing in some region of space for the integral to be nonzero. A similar argument applies to j and b, of course. This alone leads to quadratic scaling of the number of significant ERIs. Furthermore, multipole expansion of $(ia \mid jb)$ reveals an asymptotic decay rate proportional to R^{-3}, where R is a measure of the distance between i and j, leading to linear scaling of the number of significant integrals [34, 35]. The amplitudes and the residuals in Equation (3) inherit the decay property of the integrals, paving the way for a linear-scaling algorithm. By the same token, it follows that the correlation-energy contribution from pairs of occupied orbitals decays asymptotically as R^{-6}, consistent with the decay of London dispersion forces. This argument can be extended to cover also higher-order CC models.

The arguments underpinning linear-scaling correlation treatments thus rely heavily on the concept of orbital locality. As mentioned in the Introduction, this concept is not uniquely defined and a number of localisation functionals have been proposed. In this work, we will consider the central-moment functionals of Høyvik and Jørgensenalong with their statistics-based measures of orbital locality [4]. The mth power of the second central

moment (PSM-m) functionals are defined in terms of the second moment orbital spread of each orbital p,

$$\sigma_2(p) = \langle p| \left(\mathbf{r} - \langle p|\mathbf{r}|p \rangle \right)^2 |p \rangle^{1/2}, \tag{4}$$

as

$$\xi^{\text{PSM}-m} = \sum_p \sigma_2(p)^{2m}, \tag{5}$$

where the summation over orbitals should be restricted to either occupied or virtual orbitals to maintain the Brillouin condition. The PSM-1 functional is identical to the Foster-Boys functional [36, 37]. Minimising the PSM-1 functional with respect to unitary rotations of the (orthonormal) orbitals leads to the set of orbitals with the smallest possible sum of orbital spreads. In the context of periodic systems, such orbitals are commonly referred to as maximally localised Wannier functions [38].

Similarly, the PSM-2 orbitals are computed by minimising the objective function in Equation (5) with $m = 2$. The PSM-2 objective function reduces the spread of the least local orbitals at the expense of the most local ones. The motivation behind the PSM-2 functional is the conjecture that the least local orbitals in the PSM-1 set lead to excessive computational effort in a local correlation treatment [22]. Increasing the value of m does not bring any further advantages with respect to the least local orbital(s).

The PSM-m objective functions do not address the problem of long-range tails of the orbitals. The tail of orbital p can be measured by the fourth moment orbital spread σ_4:

$$\sigma_4(p) = \langle p| \left(\mathbf{r} - \langle p|\mathbf{r}|p \rangle \right)^4 |p \rangle^{1/4}, \tag{6}$$

and letting this quantity take the role of $\sigma_2(p)$ leads to the mth power of the fourth central moment (PFM-m) class of localisation functionals [4, 16]

$$\xi^{\text{PFM}-m} = \sum_p \sigma_4(p)^{4m}. \tag{7}$$

Minimising the PFM-1 objective function leads to 'minimally tailed' orthonormal orbitals and, in analogy with the PSM-2 case above, heavy-tailed outliers may be removed by putting $m = 2$. Following Høyvik and Jørgensen [4], we use the tail spread $\beta(p)$, defined as the fourth root of the kurtosis,

$$\beta(p) = \frac{\sigma_4(p)}{\sigma_2(p)}, \tag{8}$$

to measure tail thickness. The more heavy-tailed an orbital is, the greater the value of its tail spread.

In this work, we will focus on orthonormal orbitals obtained by minimising the PSM-1, PSM-2, and PFM-2 objective functions.

The by far most commonly used virtual basis is composed of PAOs [8], which are straightforwardly constructed by projecting out the occupied orbitals from the atomic orbital (AO) basis, so that a normalised PAO $|\tilde{\mu}\rangle$ is given by

$$|\tilde{\mu}\rangle = N_\mu \left(1 - \sum_i |i\rangle\langle i| \right) |\mu\rangle, \qquad (9)$$

where $|\mu\rangle$ are the AOs and $N_\mu = \langle\mu|(1 - \sum_i |i\rangle\langle i|)|\mu\rangle^{-1/2}$. In this way, a redundant set of orbitals is obtained, having in principle the same size as the AO basis. The PAOs inherit a certain degree of locality from the HF density matrix used for the projection and are free from the delocalisation (tail) effects resulting from the orthogonality constraint [24].

We are in this work mainly interested in the impact of the choice of virtual orbitals on the efficiency and accuracy of periodic local MP2 calculations. Although the choice of virtual orbitals can not be entirely decoupled from the choice of localised occupied orbitals, we will use the PSM-1 localised occupied orbitals in conjunction with different sets of either LVOs or PAOs. As proposed by Pinski et al. [30], we may then characterise the so-called multiplicative sparsity between the occupied and virtual sets of orbitals through differential overlap integrals (DOIs). The DOI for an occupied orbital φ_i and a virtual orbital φ_a is defined as

$$\Omega_{ia} = \sqrt{\int \varphi_i(\boldsymbol{r})^2 \varphi_a(\boldsymbol{r})^2 \, d\boldsymbol{r}}. \qquad (10)$$

The DOIs measure the extent to which the occupied and virtual orbitals have a non-vanishing intersection in $\mathbb{R}^3$ and thus can be used to map the sparsity of the ERIs, Equation (2), with the chosen virtual orbital basis.

While the focus of the localising objective functions and construction procedures is on the spatial distribution of the orbitals, they are each implicitly assumed to bring forth a condition where the significant part of the correlation can be attributed to a subset of the full amplitude space, determined from locality considerations. Although most modern implementations of local correlation theories reduce the impact of the initial choice of virtual basis through further refinements like orbital-specific virtuals [39] or pair-natural orbitals [30, 32, 33], it is of fundamental importance to compare the behaviour of PAOs and orthonormal virtual orbitals in the context of computing the ground-state correlation energy. It is, however, not trivial to devise a fair comparison between different types of orbitals (orthonormal or non-orthonormal, atom-centred or not), where the correlation energy is closely related to the number of virtual orbitals used in the calculation. In this work, we have chosen to treat the excitation space in the following manner:

(1) For each occupied orbital a list of atomic centres in its vicinity is compiled, and PAOs centred on such atoms constitute the initial excitation domain. This list is built by taking the n atoms closest to the occupied-orbital centroid, and then rounding up such that all atoms within the same distance are included (within a numerical threshold).
(2) Through the orbital-specific virtuals (OSV) methodology [39], a set of virtuals is built as a linear combination of the initial set by diagonalising the MP2 pair density matrices $\mathbf{D}^{ii}$ corresponding to diagonal pairs ii in the reference unit cell:

$$D_{ab}^{ii} = \sum_c t_{ii}^{ac} t_{ii}^{cb} \qquad (11)$$

(3) The resulting set of orbitals is trimmed to keep the error in the total correlation energy within a threshold.
(4) The number of OSVs retained for each domain is then a function of (i) the size of the original PAO domain, and (ii) the number of orbitals significantly contributing to the correlation energy.
(5) Finally, the virtual space of a local pair domain (which, for diagonal pairs, coincides with the orbital domain) is transformed into a local orthonormal (LON) space for the pair-domain amplitudes, in which redundant orbitals are discarded according to a threshold of 10^{-4} on the overlap eigenvalue.

While thresholds on the LON redundancy check in point (5) above are generally tighter in molecular applications, the closer packing of bulk solids makes the algorithm more sensitive to linear dependencies and, by experience, the chosen value is reasonable. Sets of LON orbitals are constructed in the same manner starting from each initial set of LVOs (PSM-1, PSM-2, or PFM-2 LVOs). In this procedure, each LVO is associated with the atom nearest to its orbital centroid. The convergence of the local MP2 correlation energy toward the exact (canonical) result as a function of the average number of LON orbitals retained in the diagonal domains provides a measure of the ability of the initial virtual-orbital (PAO or LVO) set to capture the main correlation effects.

3. Computational details

We shall consider MP2 energy calculations for three simple bulk insulators: diamond, LiH, and HCN. These are chosen as representatives of the different chemical situations that can be found in a nonmetallic (and non-magnetic) solid: a purely covalently bonded, a purely ionic, and a molecular crystal. The systems are also chosen to keep the complexity low enough to facilitate analysis of the orbitals and their effect on correlation energies.

The localised occupied orbitals are obtained from the Wannierization algorithm [40] implemented in the CRYSTAL program [41]. This procedure includes a Foster-Boys localisation [11, 12] that minimises the PSM-1 objective function given in Equation (5). The PAOs are constructed as in Equation (9) and those with norm below 10^{-3} are discarded.

As opposed to many other codes developed for ab initio studies of solid-state materials, CRYSTAL adopts atom-centred Gaussian basis sets, making it comparable with molecular programs – and, in fact, fully equivalent when zero–dimensional periodicity is invoked. A 6-31G** basis was used for HCN [42, 43], 6-21G* for diamond [44], while a simpler basis set was adopted for LiH (3-11G** on H [45], 6-1G** on Li [46]). Detailed structures (nuclear coordinates, unit cell parameters) and technical details regarding the initial HF calculations are provided in the Appendix. In all cases the experimental lattice parameters have ben adopted. The lattice constants of the cubic diamond and LiH crystal structures are 3.56679 Å [47] and 4.0834 Å [48], respectively, while those of the tetragonal HCN crystal structure are $a = 4.13$ Å, $b = 4.85$ Å, and $c = 4.34$ Å [49]. Note that we use primitive unit cells in all calculations.

The LVO orbitals are generated by the same Wannierization algorithm in CRYSTAL, yielding the PSM-1 LVOs (see the Appendix for more technical details). The PSM-2 and PFM-2 LVOs are obtained by further optimisation of the PSM-1 LVOs in a stochastic procedure where only
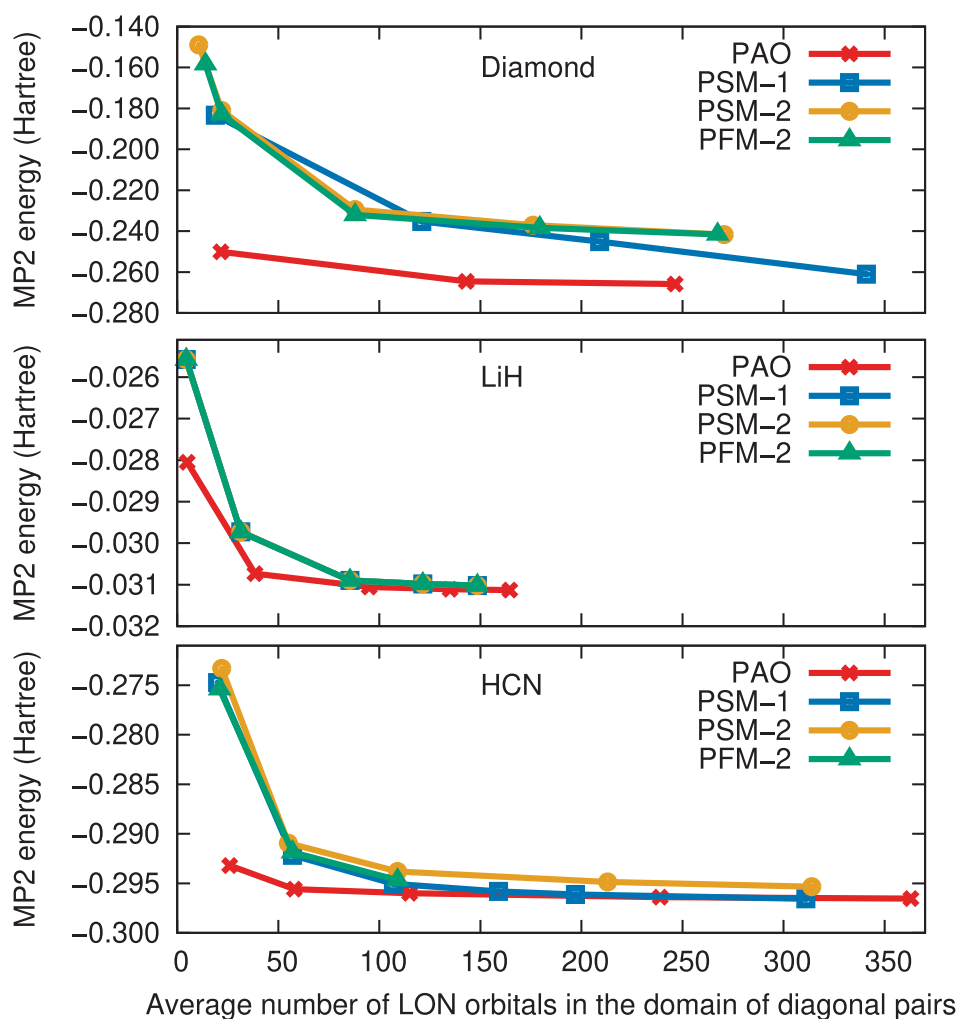

Figure 1. Convergence of the MP2 correlation energy with respect to the average number of LON orbitals in the diagonal pairs. The panels from top to bottom show results for bulk diamond, LiH and HCN, respectively, for the PAO and various LVO sets.

unitary transformations amongst the virtual orbitals in the reference unit cell are permitted. The LVOs associated with all other unit cells of the crystal are then obtained by simple translations.

For each representation of the virtual space, we quantify locality by computing second- and fourth moment orbital spreads, Equations (4) and (6), as well as the tail spread, Equation (8). Further analysis of sparsity in the orbital sets is performed by estimating the DOIs, Equation (10), using the Monte Carlo integration scheme outlined in appendix II.

Periodic local MP2 calculations are performed with the CRYSCOR suite [50–54]. Within the present CRYSCOR implementation [54], the initial excitation domains are constructed by means of the OSV method as outlined in the previous section with the default number of neighbouring atoms $n = 30$. This number is incrementally increased for each of the systems under consideration, recording the local MP2 correlation energy as a function of the average number of LON orbitals in the domains of diagonal pairs. As in Refs. [17, 54], symmetric OSV pair domains are used in the evaluation of the MP2 correlation energy.

4. Results and discussion

Figure 1 shows the convergence behaviour of the local MP2 correlation energy with respect to the average size of the LON space for each set of initial virtual orbitals. The convergence is strikingly better with PAOs than with LVOs for all three systems. Across a range of chemically distinct systems – covalent diamond, ionic LiH and molecular HCN – we thus observe that the PAOs capture more of the correlation energy within smaller domains than the PSM-1, PSM-2, and PFM-2 LVOs. Overall, the performance of the three LVO sets is equally poor. For LiH, the PAO set yields an essentially converged correlation energy with an average LON dimension of about 40, whereas roughly twice as many LON orbitals are required with the LVO sets. For HCN, the LON dimension required for convergence with the LVO sets is roughly thrice that of the PAO set. The diamond crystal is more challenging, also with PAOs, which yield convergence at a LON dimension just below 150, wheras LVOs require at least about 350 LON orbitals.

As apparent from the locality measures compiled in Figure 2, the superior convergence of the PAOs can not be unambiguously attributed to their being more local or light-tailed than the LVOs. The σ_2 data show that the PAOs are on average the least local choice and, consequently, orbital spreads can not explain the very different convergence behaviours. This agrees with the observations for molecules by Krause and Werner [25]. Nor
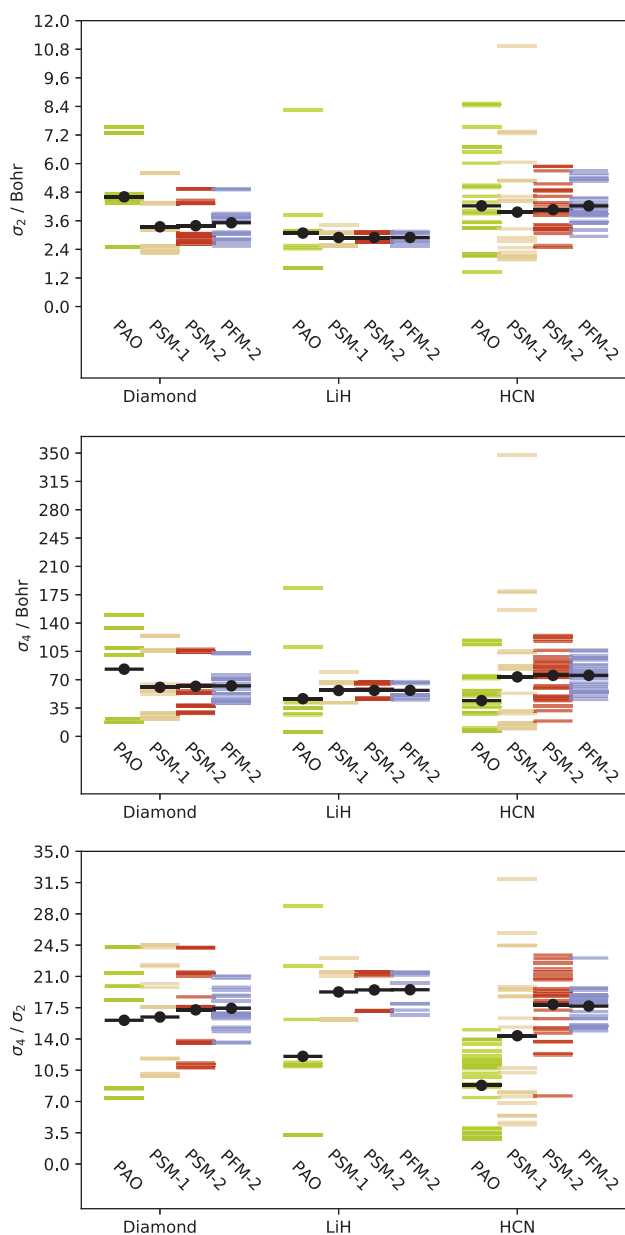

Figure 2. Second- and fourth moment orbital spreads and tail spreads for the various sets of virtual orbitals for diamond, LiH, and HCN. Horizontal lines give the value for each orbital in each set, while average values are indicated by the horizontal black lines.

can a clear distinction of the PAOs be made from the σ_4 data. While the PAOs do have the smallest average σ_4 for LiH and HCN, they also show the greatest maximum value for LiH and diamond. In all cases, the smallest maximum σ_4 is observed for the PFM-2 set. The average β values, on the other hand, are smaller for the PAOs than for the LVO sets across all three systems. While this is fully consistent with the observed correlation-energy convergence behaviour in Figure 1, we note that the maximum β value is greater for the PAOs than for the PFM-2 sets for both diamond and LiH. Judging from the β values alone, one would expect the
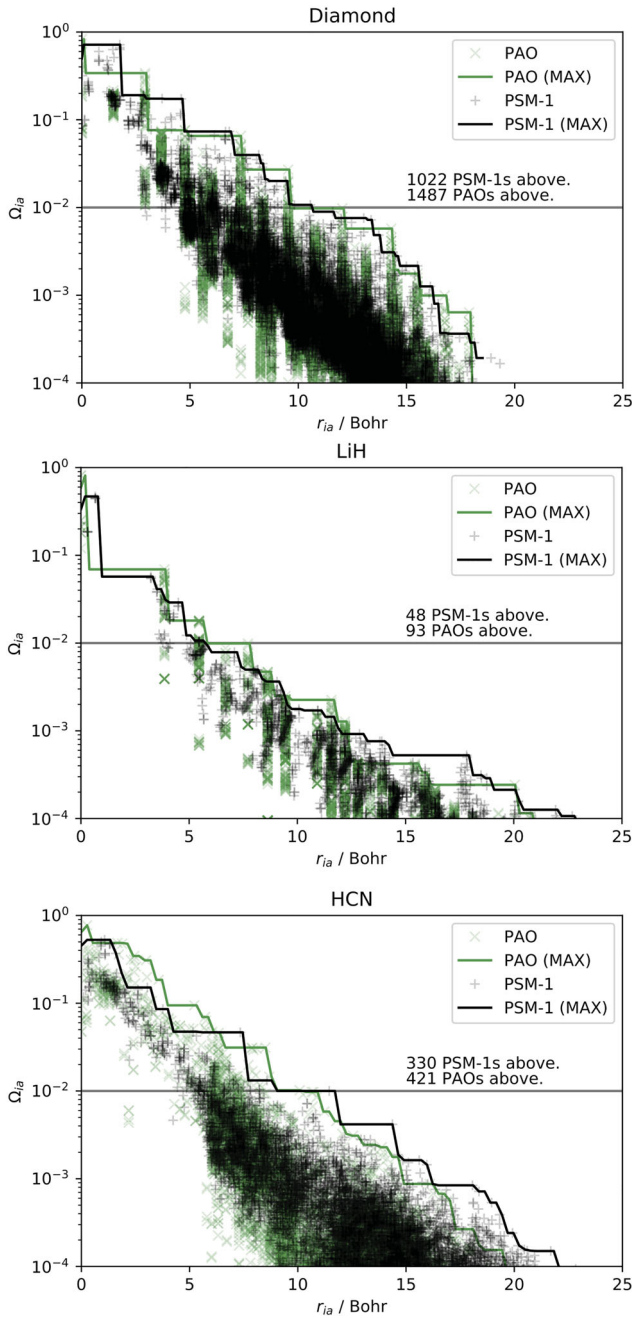
Figure 3. Differential overlap integrals between the occupied orbitals of the reference unit cell and the PAO and PSM-1 virtual orbital spaces for diamond (top), LiH (center), and HCN (bottom), as a function of the inter-orbital distance r_{ia} between occupied orbitals in the reference cell and PAOs or PSM-1 virtual orbitals throughout the crystal. The graphs show the maximum DOI for each orbital set above a given distance.

correlation-energy convergence behaviour to be similar for PAOs and either of the LVO sets in the case of diamond, while the PAOs should be vastly superior to the LVOs in the case of HCN. As seen from Figure 1, however, the LVO sets struggle even more for diamond than for HCN and LiH. We thus conclude that neither second moment spread, fourth moment spread, nor tail spread

can unambiguously explain the observed convergence behaviour.

An alternative explanation might be that the PAOs provide greater multiplicative sparsity, translating into smaller diagonal excitation domains by allowing for a smaller cutoff distance from each occupied orbital. According to this hypothesis, the PAOs should show few important DOIs at short range followed by rapid decay at longer distances, whereas the LVOs should decay more slowly with distance. The estimated DOIs for PAOs and PSM-1 virtual orbitals are plotted in Figure 3. For each crystal, the overall DOI decay rates are comparable for both sets of virtual orbitals, although the DOIs of the PAOs are more scattered with values at each distance both above and below those of the PSM-1 orbitals. A DOI cutoff threshold of 10^{-2}, as recommended by Pinski et al. [30], would suggest that about the same or even *more* PAOs than PSM-1 orbitals must be included in the correlation treatment.

Evidently, neither the statistics-based locality measures nor the DOIs provide a clear-cut explanation of the more rapid correlation-energy convergence with the PAOs. Although the LVOs span the same space as the PAOs when the entire crystal is considered, a distance-based truncation of the PAOs clearly captures more of the correlation energy. From the viewpoint of the LVOs, this implies that the PAOs of a given cell must contain components that are only present in more distant orbitals in the orthonormal representation. To investigate this effect, we expand the PAOs of the reference unit cell, denoted $|0\tilde{\mu}\rangle$ where 0 indicates that the parent AO $|\mu\rangle$ is located in the reference unit cell, in a basis defined by a given set of LVOs,

$$|0\tilde{\mu}\rangle = \sum_L \sum_a |La\rangle\langle La \mid 0\tilde{\mu}\rangle \equiv \sum_L \sum_a |La\rangle C_{a\tilde{\mu}}^L, \quad (12)$$

where L is a lattice index (including 0, the reference unit cell). The expansion coefficients can be computed directly from their definition. In order to measure the extent to which distant LVOs are represented in the PAO set of the reference unit cell, we introduce the 'weight' of each virtual orbital $|La\rangle$ as the square of the corresponding expansion coefficient,

$$W_{a\tilde{\mu}}^L = \left(C_{a\tilde{\mu}}^L\right)^2. \quad (13)$$

These weights are plotted in Figure 4 as a function of the inter-orbital distance $r_{\tilde{\mu}a}$ between the PAOs in the reference cell and PSM-1 LVOs throughout the crystal.

Only plots for the PSM-1 set are shown. Essentially identical plots are obtained for the PSM-2 and PFM-2 sets. While the weights decay quite rapidly, significant contributions ($W_{a\tilde{\mu}}^L \gtrsim 10^{-3}$) may indeed be
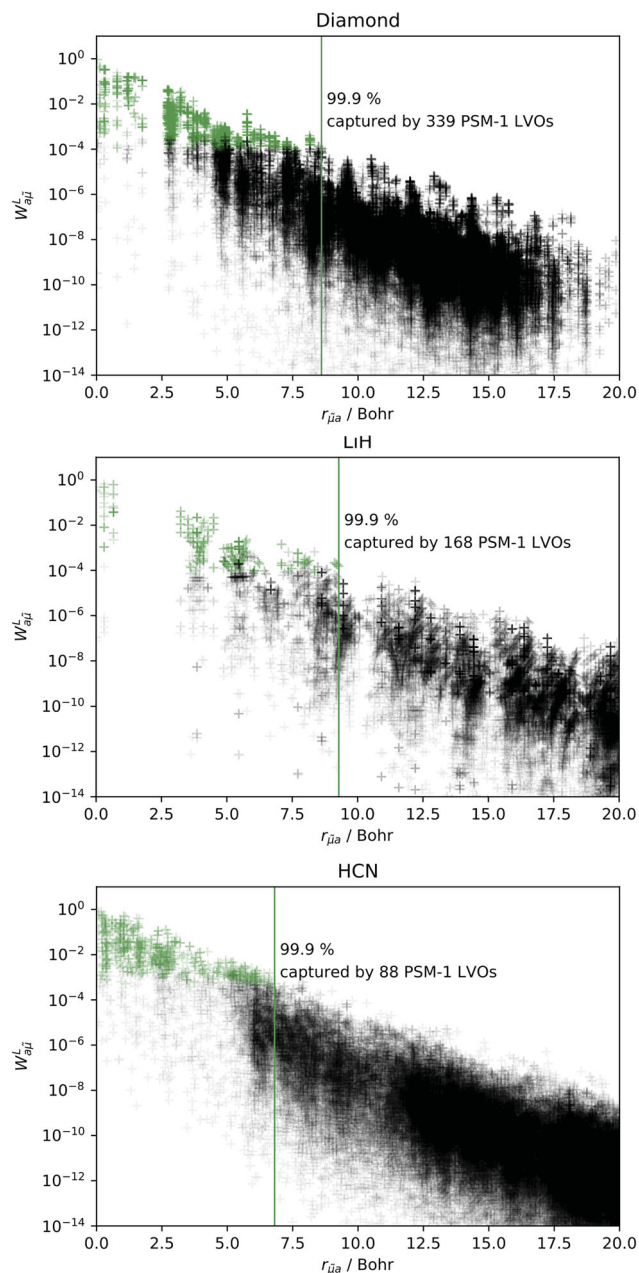
Figure 4. Expansion coefficients squared, Equation (13), of the PSM-1 virtual orbitals for the reference-cell PAOs as a function of the distance between the reference cell PAOs and each virtual orbital for diamond (top panel), LiH (middle panel), and HCN (bottom panel). Green points indicate expansion coefficients that are required to recover 99.9% of the norm of every single PAO. There are 26, 11 and 32 such PAOs for diamond, LiH, and HCN, respectively. The most distant orbital that has to be included to capture the same span is indicated with a vertical green line.

observed between 5 and 10 bohr from the origin of the reference cell. This is consistent with the observed correlation-energy convergence behaviour. Furthermore, it is evident from the vertical lines of Figure 4 that a distance-based truncation of the LVOs leads to the inclusion of a large number of redundant components, ultimately resulting in inefficient (but still linear-scaling)

calculations of the MP2 correlation energy. Interestingly, the distribution of the green points suggests that a more sparse subspace of LVOs may be chosen based on their importance in spanning the same space as the reference-cell PAOs.

5. Concluding remarks

We have studied the impact of the initial choice of virtual-orbital basis, PAOs and three different sets of LVOs, on the convergence of the local MP2 correlation energy toward the canonical result for three insulating crystals with qualitatively different chemical bonding situations, diamond (covalent crystal), LiH (ionic crystal), and HCN (molecular crystal). Our results confirm recent findings by Werner and coworkers [25–27] for molecules: the performance of PAOs is significantly better than LVOs.

This result is somewhat counterintuitive since the LVOs are generally more local according to second moment orbital spread data, have thinner tails according to fourth moment orbital spread data, and provide comparable or even greater multiplicative sparsity in conjunction with localised orthonormal occupied orbitals than the PAOs. Although we do observe a certain agreement with tail spread data for the PAOs and LVOs, correctly indicating superiority of the former, there is no one-to-one mapping between β values and the observed convergence behaviour. Inspecting the expansion of the reference-cell PAOs in the LVO basis, we find that the PAOs contain surprisingly large components of distant LVOs. It seems, therefore, that the efficiency of the PAOs can be traced to their being sufficiently local *and* linearly dependent such that by choosing a subset of them based on the location of the atomic centres of the parent AOs, we get a greater fraction of the excitation space than we would have obtained with the LVOs centred in the same spatial region.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

A. S. Hansen 🔘 http://orcid.org/0000-0003-0962-3143
E. Rebolini 🔘 http://orcid.org/0000-0001-5709-5547
L. Maschio 🔘 http://orcid.org/0000-0002-4657-9439
T. B. Pedersen 🔘 http://orcid.org/0000-0001-8967-6055

References

[1] A.E. Reed, R.B. Weinstock and F. Weinhold, J. Chem. Phys. **83**, 735–746 (1985).

[2] A.E. Reed, L.A. Curtiss and F. Weinhold, Chem. Rev. **88**, 899–926 (1988).

[3] G. Knizia, J. Chem. Theory Comput. **9**, 4834–4843 (2013).

[4] I.M. Høyvik and P. Jørgensen, Chem. Rev. **116**, 3306–3327 (2016).

[5] M. Ziółkowski, B. Jansik, P. Jørgensen and J. Olsen, J. Chem. Phys. **131**, 124112 (2009).

[6] J.M. Herbert, J. Chem. Phys. **151**, 170901 (2019).

[7] D. Usvyat, L. Maschio and M. Schütz, WIREs: Comp. Mol. Sci. **8**, e1357 (2018).

[8] P. Pulay, Chem. Phys. Lett. **100**, 151–154 (1983).

[9] M. Schütz, G. Hetzer and H.J. Werner, J. Chem. Phys. **111**, 5691–5705 (1999).

[10] F. Neese, A. Hansen and D.G. Liakos, J. Chem. Phys. **131**, 064103 (2009).

[11] J.M. Foster and S.F. Boys, Rev. Mod. Phys. **32**, 300–302 (1960).

[12] S.F. Boys, Rev. Mod. Phys. **32**, 296–299 (1960).

[13] J. Pipek and P.G. Mezey, J. Chem. Phys. **90**, 4916 (1989).

[14] C. Edmiston and K. Ruedenberg, Rev. Mod. Phys. **35**, 457 (1963).

[15] C. Edmiston and K. Ruedenberg, J. Chem. Phys. **43**, S97 (1965).

[16] I.M. Høyvik, B. Jansik and P. Jørgensen, J. Chem. Phys. **137**, 224114 (2012).

[17] J. Yang, G.K.L. Chan, F.R. Manby, M. Schütz and H.J. Werner, J. Chem. Phys. **136**, 144105 (2012).

[18] R. Ahlrichs and W. Kutzelnigg, J. Chem. Phys. **48**, 1819–1832 (1968).

[19] S. Liu, J.M. Pérez-Jordá and W. Yang, J. Chem. Phys. **112**, 1634–1644 (2000).

[20] B. Paulus, K. Rościszewski, H. Stoll and U. Birkenheuer, Phys. Chem. Chem. Phys. **5**, 5523–5529 (2003).

[21] D. Usvyat and M. Schütz, Theor. Chem. Acc. **114**, 276–282 (2005).

[22] B. Jansík, S. Høst, K. Kristensen and P. Jørgensen, J. Chem. Phys. **134**, 194104 (2011).

[23] I.M. Høyvik, K. Kristensen, B. Jansik and P. Jørgensen, J. Chem. Phys. **136**, 014105 (2012).

[24] I.M. Høyvik, K. Kristensen, T. Kjærgaard and P. Jørgensen, Theor. Chem. Acc. **133**, 1–10 (2014).

[25] C. Krause and H.J. Werner, Phys. Chem. Chem. Phys. **14**, 7591–7604 (2012).

[26] H.J. Werner, C. Köppl, Q. Ma and M. Schwilk, in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, edited by Mark S. Gordon (John Wiley & Sons, Ltd, Chichester, UK, 2017), Chap. 1, pp. 1–79.

[27] Q. Ma and H.J. Werner, WIREs: Comp. Mol. Sci. **8**, e1371 (2018).

[28] T. Helgaker, P. Jørgensen and J. Olsen, *Molecular Electronic-Structure Theory* (John Wiley and Sons, Ltd, Chichester, 2000).

[29] Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki and M. Kállay, J. Chem. Phys. **139**, 094105 (2013).

[30] P. Pinski, C. Riplinger, E.F. Valeev and F. Neese, J. Chem. Phys. **143**, 034108 (2015).

[31] J.J. Eriksen, P. Baudin, P. Ettenhuber, K. Kristensen, T. Kjærgaard and P. Jørgensen, J. Chem. Theory Comput. **11**, 2984–2993 (2015).

[32] C. Riplinger, P. Pinski, U. Becker, E.F. Valeev and F. Neese, J. Chem. Phys. **144**, 024109 (2016).

[33] H.J. Werner, G. Knizia, C. Krause, M. Schwilk and M. Dornbach, J. Chem. Theory Comput. **11**, 484–507 (2015).

[34] G. Hetzer, P. Pulay and H.J. Werner, Chem. Phys. Lett. **290**, 143–149 (1998).

[35] H.J. Werner, J. Chem. Phys. **145**, 201101 (2016).

[36] S.F. Boys, Rev. Mod. Phys. **32**, 296 (1960).

[37] J.M. Foster and S.F. Boys, Rev. Mod. Phys. **32**, 300 (1960).

[38] N. Marzari, A.A. Mostofi, J.R. Yates, I. Souza and D. Vanderbilt, Rev. Mod. Phys. **84**, 1419–1475 (2012).

[39] J. Yang, Y. Kurashige, F.R. Manby and G.K.L. Chan, J. Chem. Phys. **134**, 044123 (2011).

[40] C.M. Zicovich-Wilson, R. Dovesi and V.R. Saunders, J. Chem. Phys. **115**, 9708–9719 (2001).

[41] R. Dovesi, A. Erba, R. Orlando, C.M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro and B. Kirtman, WIREs: Comp. Mol. Sci. **8**, e1360 (2018).

[42] R. Dovesi, C. Ermondi, E. Ferrero, C. Pisani and C. Roetti, Phys. Rev. B **29**, 3591–3600 (1984).

[43] C. Gatti, V.R. Saunders and C. Roetti, J. Chem. Phys. **101**, 10686–10696 (1994).

[44] M. Catti, A. Pavese, R. Dovesi and V.R. Saunders, Phys. Rev. B **47**, 9189–9198 (1993).

[45] M.F. Peintinger, D.V. Oliveira and T. Bredow, J. Comput. Chem. **34**, 451–459 (2013).

[46] B. Civalleri, A.M. Ferrari, M. Llunell, R. Orlando, M. Mérawa and P. Ugliengo, Chem. Mat. **15**, 3996–4004 (2003).

[47] R.W.G. Wyckoff, *Crystal Structures*, Vol. 1, 2nd ed. (Interscience Publishers, New York, 1963).

[48] E. Staritzky and D. Walker, J. Annal. Chem. **28**, 1055 (1956).

[49] W.J. Dulmage and W.N. Lipscomb, Acta Cryst. **4**, 330 (1951).

[50] C. Pisani, M. Busso, G. Capecchi, S. Casassa, R. Dovesi, L. Maschio, C. Zicovich-Wilson and M. Schütz, J. Chem. Phys. **122**, 094113 (2005).

[51] L. Maschio, D. Usvyat, F. Manby, S. Casassa, C. Pisani and M. Schütz, Phys. Rev. B **76**, 075101 (2007).

[52] C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz and D. Usvyat, J. Comput. Chem. **29**, 2113–2124 (2008).

[53] C. Pisani, M. Schütz, S. Casassa, D. Usvyat, L. Maschio, M. Lorenz, A. Erba, M. Schutz, S. Casassa, D. Usvyat, L. Maschio, M. Lorenz and A. Erba, Phys. Chem. Chem. Phys. **14**, 7615–7628 (2012).

[54] D. Usvyat, L. Maschio and M. Schütz, J. Chem. Phys. **143**, 102805 (2015).

[55] Crystal manual http://www.crystal.unito.it/documentation.php 2019 (Accessed: 2019-12-03).

[56] S.T. Tokdar and R.E. Kass, WIREs Comp. Stat. **2**, 54–60 (2010).

Appendices

Appendix 1. Geometries

The reader is referred to the Crystal manual [55] for details regarding the keywords given in the following discussion. All calculations was performed for 3D periodicity. In all cases the two-electron integrals was treated exactly (invoked with keyword NOBIPOLA) when setting up the Fock-matrix.

Diamond was computed with space group 227 and lattice parameter 3.56679 Å. With a standard shift of the origin (IFSO set to 1), a carbon was placed in $r_C = (0, 0, 0)$. The Hartree-Fock optimisation was performed with SHRINK set to 8 and an energy tolerance of 10^{-10}. The intergral tolerances was 8 8 8 10 20. For the Wannierization, we used a NEWK of 9.

For LiH we used space group 225 with lattice parameter 4.0834 Å. Lithium was placed at $r_{Li} = (0, 0, 0)$ and hydrogen at $r_H = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, both in units given as fractions of the lattice vectors. For the Hartree-Fock optimisation we used a SHRINK factor of 7 with a convergence tolerance on the energy (TOLDEE) 10^{-12}. Integral tolerances (ITOL) was set to 8 8 8 25 50. For the Wannierization we used a NEWK of 11.

For HCN we used space group 44 and lattice parameters 4.13 Å, 4.85 Å, 4.34 Å, 90°, 90°, and 90°. The fractional coordinates of hydrogen, carbon and nitrogen were $r_H = (0, 0, -0.2459793977425)$, $r_C = (0, 0, 0.003572703972826)$, and $r_N = (0, 0, 0.2701066937697)$, respectively. For the Hartree-Fock optimisation we used a SHRINK factor of 7 and integral tolerances of 7 7 7 20 40. The energy tolerance (TOLDEE) was set to 10^{-10}. The subsequent Wannierization was performed with NEWK 11.

Appendix 2. DOI estimates

The differential overlap integrals (DOIs) [30] were estimated using Monte Carlo (MC) integration [56] within finite volumes of the full integration domain ($\mathbb{R}^3$).
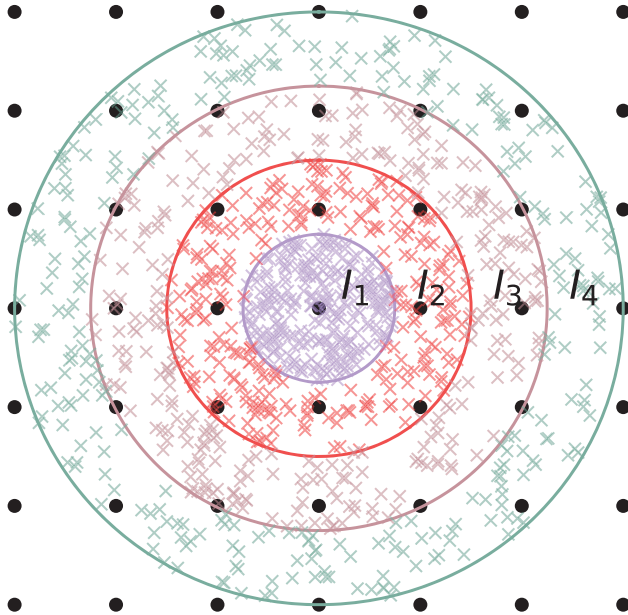

Figure A1. Illustration of Monte Carlo integration within concentric shells. The density of samples is higher in the centre. The final integral is estimated by $I = \sum_{n=1}^{N_{shells}} I_n$.
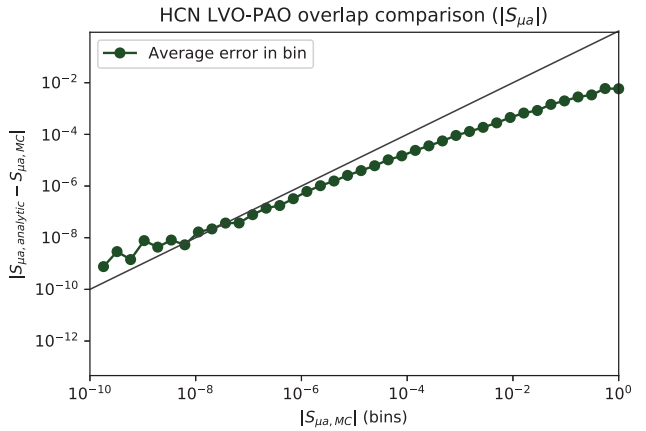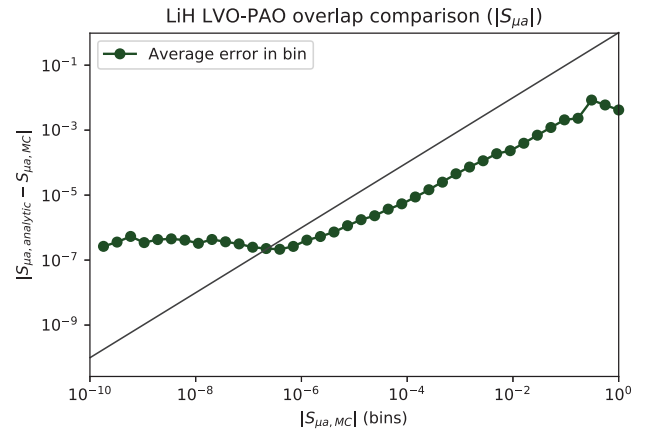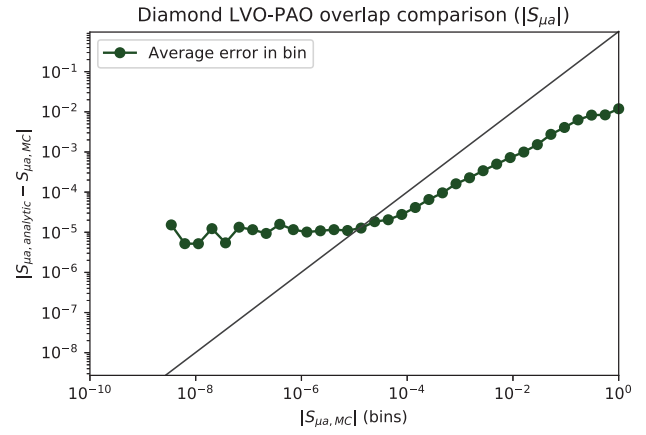

Figure A2. Absolute difference of analytical and MC estimates for the PAO-LVO ($S_{\mu a}$) overlap matrix elements. The elements are sorted into bins depending on the absolute value of the MC-estimates, whereby the average error within each bin is calculated and plotted. The diagonal line shows where the absolute deviation equals the MC-estimate. The intersection between the diagonal and the average can be used to identify a lower bound at which the loss of precision in the MC estimates becomes too severe to draw any conclusions. This intersection appears well below 10^{-4} in all three cases.

For each pair of unit cells, the full integration domain is divided into concentric spherical shells as illustrated in Figure A1, with their origin situated halfway between the lattice vectors associated with the relevant cells. The outermost layer

is placed at 50 Bohr, with its interior divided into 121 shells uniformly spaced in the radial direction.

The overlap and differential overlap integrals are computed 20 times for all orbitals associated with the two cells using MC integration with 2000 random samples uniformly distributed inside every finite volume. The integrals over every volume are then summed, and the final integrals are estimated as the average of the 20 separate estimates.

In total, each integrand is sampled in 4.84×10^6 coordinates. The discretisation of the integration domain can be considered a discrete importance sampling [56], with the consequence of more dense sampling in regions where higher variance in the integrand is expected. The assumption of the concentric spherical shells being a reasonable sample distribution thus relies on the orbitals being localised to their associated cells, so that the products of orbitals have most of their significant distribution between the cells.

Confidence in the estimates can be established by (1) assessment of the error in the PAO-LVO overlap matrix, and (2) assuming that the same error applies to the DOIs due to the similarities in the integrand with respect to variance. While we can not directly infer the error in the DOIs from the error in the PAO-LVO overlaps, it is clear that the regions with most variance in the integrands coincide in these cases. Hence, if the sampled coordinates has the ability to reproduce the overlaps they should be an equally reasonable choice for the DOIs.

The absolute deviation of the MC estimates of the PAO-LVO overlap matrix elements is compared to their magnitude in Figure A2, where we observe a significant loss of relative accuracy below 10^{-4} for diamond, and below 10^{-6} for LiH and HCN. We thus conclude that the DOI estimates presented in Figure 3 are reliable.