Divide-Expand-Consolidate Second-Order Møller-Plesset Theory with Periodic Boundary Conditions

Elisa Rebolini,* Gustav Baardsen, Audun Skau Hansen, Karl R. Leikanger, and Thomas Bondo Pedersen

Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo, Norway

E-mail: elisa.rebolini@kjemi.uio.no

Abstract

We present a generalization of the divide-expand-consolidate (DEC) framework for local coupled-cluster calculations to periodic systems and test it at the second-order Møller-Plesset (MP2) level of theory. For simple model systems with periodicity in one, two, and three dimensions, comparisons with extrapolated molecular calculations and the local MP2 implementation in the Cryscor program show that the correlation energy errors of the extended DEC (X-DEC) algorithm can be controlled through a single parameter, the fragment optimization threshold. Two computational bottlenecks are identified: the size of the virtual orbital spaces and the number of pair fragments required to achieve a given accuracy of the correlation energy. For the latter, we propose an affordable algorithm based on cubic splines interpolation of a limited number of pair-fragment interaction energies to determine a pair cutoff distance in accordance with the specified fragment optimization threshold.

1 Introduction

While coupled-cluster (CC) theory[1,2] is widely accepted as the most successful electronic-structure method in the sense that it defines a convergent hierarchy of increasingly accurate approximations to electronic ground- and excited-state energies and properties of molecular systems, it is not the most frequently used method in computational chemistry. There are two main reasons for this. First, the standard formulation of CC theory is based on a single Hartree-Fock (HF) reference determinant, which generally must be a sufficiently good approximation to the exact wave function[3], thus excluding systems with multiconfigurational (ground-) states. Second, the polynomial-scaling computational cost of CC theory limits the applicability to small molecules. Consequently, much effort has been directed towards the formulation of reduced-scaling implementations of CC methods in recent years.

The key to reduced-scaling or, ideally, linear-scaling calculations is exploitation of the locality of electron correlation, as was pointed out by Sinanoğlu[4] and Nesbet[5] more than half a century ago. Local correlation algorithms were pioneered by Pulay and Sæbø[6–10] in the 1980s, using the freedom granted by HF theory to select bases for the occupied and virtual orbital spaces that expose the proper locality of electronic excitation amplitudes. This has led to a variety of local correlation approaches, with the local CC methods by Werner and co-workers[11–14], and by Neese and co-workers[15–18] as prominent recent examples. Note that in this paper, second-order Møller-Plesset (MP2) theory is classified as the lowest-level method in the CC hierarchy: MP2, coupled-cluster singles-and-doubles (CCSD), coupled-cluster singles-and-doubles with perturbative triples (CCSD(T)), coupled-cluster singles-and-doubles-and-triples (CCSDT), and so on.

Whether local or canonical, most CC developments have been aimed at finite systems, whereas extended systems have received relatively little attention. The main reason is the steep increase in computational cost compared to Kohn-Sham density-functional theory, which is the undisputed workhorse of computational electronic-structure theory. Although periodic boundary conditions allow us to formally map the infinite problem onto a single

reference unit cell, Brillouin-zone sampling increases the computational effort by several orders of magnitude compared to the cost of a single unit cell treated as a finite system. Consequently, even for small unit cells, *exact* canonical (nonlocal) implementations of CC theory face a steep scaling wall, especially for three-dimensional (3D) systems.

For one-dimensional (1D) systems, Hirata, Bartlett and co-workers[19–22] have presented Gaussian-based canonical CCSD implementations where infinite lattice sums are truncated by prescreening. More recently, Grüneis, Kresse, Alavi *et al.*[23–26] reported plane-wave-based canonical MP2, CCSD, and CCSD(T) implementations applicable to 3D systems through virtual space truncations. Using Gaussian basis functions, McClain *et al.*[27] also reported canonical MP2 and CCSD implementations for 3D systems, which avoid truncation of the virtual space and thus are limited to modest Brillouin-zone sampling. In addition to ground-state energies, McClain *et al.*[27] implemented equation-of-motion CC[28,29] for computing correlated band structures. Ayala *et al.*[30] used Almlöf's Laplace transform technique to eliminate the orbital-energy denominators[31–33] and recast the MP2 expression in Gaussian-type atomic-orbital (AO) basis to exploit sparsity. Del Ben *et al.*[34,35] devised a massively parallel MP2 algorithm with a hybrid Gaussian and plane-wave basis and Brillouin-zone sampling restricted to the Γ point, handling instead the finite-size problem by a supercell approach.

While embedding[36–38] and incremental[39–43] techniques may also be viewed as local correlation approaches, the only implementation of a fully 3D periodic local CC method is the MP2 approximation available in the Gaussian-based CRYSCOR program[44–49]. It is essentially the molecular approach of Werner and co-workers[11] subjected to periodic boundary conditions, using (generalized) Wannier orbitals[50–52] for the occupied space and Pulay's projected atomic orbitals (PAOs)[6] for the virtual space.

Infinite periodic systems are significantly more challenging than finite molecular systems from the perspective of numerical accuracy, since several cutoffs and other approximations must be invoked to make the calculations feasible. This makes comparisons of, say, correlation energies from different algorithms difficult, not least for advanced electronic-structure

methods like CC theory, and a simple effective mechanism for controlling error is of the utmost importance. In this paper we present a preliminary investigation of the periodic generalization of the divide-expand-consolidate (DEC) framework for molecular CC methods developed by Jørgensen and co-workers[53–58]. The DEC framework supports the complete hierarchy of ground-state CC models, including analytic evaluation of forces[59], is massively parallelizable[60], and offers error control through a single parameter[61]. The single-parameter error control is particularly important for the generalization of the DEC framework to periodic systems, enabling self-validation by straightforward convergence tests.

The molecular DEC framework[53–61] is based on localized occupied *and* virtual orbitals, which are individually assigned to atomic sites by some physically motivated means such as Mulliken charge, Löwdin charge, or orbital centroid. This allows us to recast the CC correlation energy as a sum over atomic fragment energies E_A and pair-fragment interaction energies ΔE_{AB}, that is,

$$E = \sum_A E_A + \sum_{B>A} \Delta E_{AB} \tag{1}$$

While this expression is formally exact and offers no computational savings, reduced scaling is achieved by restricting the orbital subspace accessible for each atom A to a neighborhood of A in such a way that the error in the atomic fragment energy is less than a given threshold, the fragment optimization threshold (FOT)[53]. The fragment orbital subspaces thus obtained are then used to compute pair-fragment interaction energies ΔE_{AB}. While the size of each atomic fragment is asymptotically constant, leading to linear scaling for the atomic fragment evaluations, the overall procedure scales quadratically with the number of atoms due to the pair fragments. It can, however, be made linear-scaling by noting that ΔE_{AB} decays asymptotically as r^{-6}, with r the distance between A and B, thus allowing truncation of the pair summation[55]. Since the atomic fragment optimizations and the subsequent pair fragment calculations are independent, the algorithm is embarassingly parallel by construction—provided, of course, enough memory is available to each compute process.

Adapting the DEC framework to extended periodic systems is, in principle, straight-

forward using (generalized) occupied and virtual Wannier functions in place of localized molecular orbitals. Very little work has been done on the localization of virtual Wannier orbitals, presumably because they would be mainly useful for local correlation methods, which have not been widely pursued for extended systems. For molecules, virtual orbitals may now be routinely computed, even with diffuse AO basis sets, using suitable localization functionals and state-of-the-art optimization algorithms[62]. Although translation symmetry complicates the situation for extended systems, the fact that well localized occupied Wannier functions can be obtained for gapped systems, including semiconductors[63], makes it possible to also obtain local virtual functions. Even if existing Wannierization algorithms may not be well suited for the virtual space in all cases, an alternative procedure would be to formulate the periodic DEC scheme starting from the direct-space PAOs, whose locality is limited by the decay behavior of the density matrix. With appropriate handling of linear dependence, the PAOs might be further localized using recently developed techniques for nonorthogonal orbital localization[64]. In this work, however, we assume that virtual Wannier functions may be computed.

For extended systems, we need to compute the correlation energy per unit cell, which is easily achieved by summing only over atoms A belonging to the reference unit cell in eq. (1). The sum over atoms B must formally extend over the entire infinite lattice, however. The pair-fragment interaction energies still decay asymptotically as r^{-6} (for non-conducting systems) but the number of pairs within a sphere of radius R is proportional to R^3 for a 3D system, implying that the number of significant pairs to be computed increases dramatically as we go from a finite molecule to an infinite 3D system. Even distant pair-fragment interaction energies may thus be important for the total correlation energy[46].

In this paper, we present an adaptation of the DEC algorithm to extended systems with periodic boundary conditions—referred to as the extended DEC (X-DEC) algorithm. Using very simple model systems, we validate the X-DEC algorithm by comparing MP2 correlation energies with those obtained from CRYSCOR and extrapolated cluster calculations. Although

computational performance is not a major concern in the present work, we analyze atomic fragment sizes from a performance perspective and investigate a procedure to determine the pair cutoff distance in a black-box manner using interpolation.

2 The extended Divide-Expand-Consolidate (X-DEC) algorithm for periodic systems

In this section, the periodic formulation of the DEC algorithm is detailed for the MP2 approximation. The X-DEC algorithm follows closely the DEC approach originally developed in the molecular framework[53–55]. For the sake of clarity, the original DEC algorithm will be referred to as "molecular DEC" in order to distinguish it from its periodic counterpart.

A sketch of the main steps in the X-DEC method is given in Algorithm 1. Setup of a reference state and orbitals is explained in Section 2.1, periodic amplitude and energy MP2 equations are presented in Section 2.2, and the specific procedures required by the periodic boundary conditions in the different steps are detailed in Sections 2.3, 2.4, and 2.5.

2.1 Local Wannier representation of the periodic Hartree-Fock wave function

The starting point for the X-DEC-MP2 calculation is a HF reference state, which is computed by the periodic code CRYSTAL[65]. This reference state is expressed in terms of intrinsically delocalized Bloch orbitals and is expanded in an atom-centered Gaussian basis set. As mentioned above, in X-DEC, both the occupied and virtual spaces must be constructed from localized orbitals. The X-DEC computations presented in this paper use Wannier functions (WFs) that are obtained from the HF Bloch orbitals by the localization-wannierization procedure implemented in the CRYSTAL program.[52] Each Wannier function $W_{p\mathbf{L}}$ is characterized by an orbital index p and a cell index $\mathbf{L}$, and orbitals outside the cell $\mathbf{0}$ are obtained by a

Algorithm 1 Outline of the X-DEC algorithm. The main steps are essentially as in the molecular DEC algorithm[53,54], but in X-DEC, atomic fragments are optimized only within the unit cell and fragments outside the unit cell are obtained by translation. Below, atoms are denoted by $\mathbf{A} \equiv (A, \mathbf{L}_A)$, where A is an atom ID and $\mathbf{L}_A$ is a cell coordinate vector.

for each atom $\mathbf{A} \equiv (A, \mathbf{0})$ **do**
    Obtain an optimized atomic fragment $F_{(A,\mathbf{0})}$.
    Compute the correlation energy $E_{(A,\mathbf{0})}$ of $F_{(A,\mathbf{0})}$.
    Store $F_{(A,\mathbf{0})}$ and $E_{(A,\mathbf{0})}$.
end for

for each atom $\mathbf{A} \equiv (A, \mathbf{0})$ **do**
    $F_{\mathbf{A}} \leftarrow F_{(A,\mathbf{0})}$
    Set up a list N_A of all neighbor atoms within a chosen cutoff radius.
    for each atom $\mathbf{B} = (B, \mathbf{L}_B)$ in the neighbor list N_A **do**
        if $\mathbf{L}_B == \mathbf{0}$ **then**
            $F_{\mathbf{B}} \leftarrow F_{(B,\mathbf{0})}$ ▷ Fragment of atom $(B, \mathbf{0})$
        else
            $F_{\mathbf{B}} \leftarrow T_{\mathbf{L}_B} F_{(B,\mathbf{0})}$ ▷ Translate fragment by $\mathbf{L}_B$.
        end if
        Set up the pair fragment $F_{\mathbf{A},\mathbf{B}}$ using $F_{\mathbf{A}}$ and $F_{\mathbf{B}}$.
        Compute and store the correlation energy $E_{\mathbf{A},\mathbf{B}}$ of $F_{\mathbf{A},\mathbf{B}}$.
    end for
end for
Sum the atomic and pair fragment energies as in Equation (9).

translation

$$W_{p\mathbf{L}}(\mathbf{r}) = W_{p\mathbf{0}}(\mathbf{r} - \mathbf{R_L}), \tag{2}$$

where $\mathbf{R_L}$ is the distance vector between the reference unit cell and the cell with coordinates $\mathbf{L}$. The Wannier orbitals satisfy the orthonormality condition

$$\int W_{p\mathbf{L}}(\mathbf{r}) W_{q\mathbf{M}}(\mathbf{r}) d\mathbf{r} = \delta_{pq}\delta_{\mathbf{LM}}, \tag{3}$$

where δ is the Kronecker delta. CRYSTAL uses Boys localization,[66] in which the sum of the orbital second central moment (PSM-1)

$$\xi_1^{\mathrm{SM}} = \sum_p \mu_2^p = \sum_p \langle p\mathbf{0}|(\mathbf{r} - \langle p\mathbf{0}|\mathbf{r}|p\mathbf{0}\rangle)^2|p\mathbf{0}\rangle \tag{4}$$

is minimized, to construct the Wannier functions. Thus obtained, the Wannier functions are often referred to as 'maximally localized'[51] in the sense of smallest possible orbital spread, $\sigma^p = \sqrt{\mu_2^p}$. For the purpose of the X-DEC algorithm, the set of Wannier orbitals obtained from CRYSTAL are then further localized by minimizing the second power of the second central moment (PSM-2)[62]

$$\xi_2^{\mathrm{SM}} = \sum_p (\mu_2^p)^2, \tag{5}$$

with respect to rotations among the occupied or the virtual orbitals in the referene unit cell only. The PSM-2 functional penalizes outliers with relatively large orbital spread at the expense of the most local orbitals, which become somewhat less local than with PSM-1.[67]

As part of the X-DEC calculation, a set of occupied and virtual Wannier functions is assigned to each atomic center in the reference unit cell. Orbitals in other cells are obtained by applying the translation operation (2). Details on the assignment of Wannier functions

to atoms can be found in Section 2.3.

Assuming a basis of real Wannier functions $\{W_{p\mathbf{L}}\}$, the two-electron integrals over these orbitals are denoted

$$(\mathbf{pq}|\mathbf{rs}) \equiv \int d\mathbf{r}_1 \int d\mathbf{r}_2 W_{\mathbf{p}}(\mathbf{r}_1) W_{\mathbf{q}}(\mathbf{r}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} W_{\mathbf{r}}(\mathbf{r}_2) W_{\mathbf{s}}(\mathbf{r}_2), \tag{6}$$

where the composite index $\mathbf{p} \equiv (p, \mathbf{L})$ is introduced for the sake of compactness.

2.2 MP2 energy and amplitude equations

In the spin-restricted formalism with a set of orthonormal HF Wannier orbitals, the MP2 correlation energy per unit cell may be written as

$$E = \sum_{\mathbf{ijab}} E_{\mathbf{ij}}^{\mathbf{ab}}, \tag{7}$$

where $\mathbf{i}$ runs over the occupied orbitals in the reference unit cell, $\mathbf{j}$ over all occupied orbitals, and $\mathbf{a}$ and $\mathbf{b}$ over all virtual orbitals. In eq. (7),

$$E_{\mathbf{ij}}^{\mathbf{ab}} = t_{\mathbf{ij}}^{\mathbf{ab}} \left\{ 2\,(\mathbf{ia}|\mathbf{jb}) - (\mathbf{ib}|\mathbf{ja}) \right\} \tag{8}$$

is the contribution to the MP2 energy coming from the orbitals $\mathbf{i}$, $\mathbf{j}$, $\mathbf{a}$, and $\mathbf{b}$, and $t_{\mathbf{ij}}^{\mathbf{ab}}$ are the MP2 amplitudes.

As in the molecular case[53–55], the summation over the orbitals $\mathbf{i}$, $\mathbf{j}$, $\mathbf{a}$, and $\mathbf{b}$ can be re-organized in terms of (on-site) atomic and pair-fragment contributions. The MP2 energy is then written as

$$E = \sum_{A} E_{(A,\mathbf{0})} + \sum_{A<B} \Delta E_{(A,\mathbf{0}),(B,\mathbf{0})} + \sum_{A} \sum_{B} \sum_{\mathbf{L}_B \neq \mathbf{0}} \Delta E_{(A,\mathbf{0}),(B,\mathbf{L}_B)}, \tag{9}$$

where the first term is a sum over atomic fragments and the two last terms represent pair-

fragment interaction energies. In an atomic fragment, $\mathbf{i}$ and $\mathbf{j}$ are assigned to the same atomic center A in the central unit cell, whereas in a pair fragment, one of the occupied orbitals is assigned to an atomic fragment A in the central unit cell and the other is associated with an atomic fragment located on atom $(B, \mathbf{L}_B)$, where $\mathbf{L}_B$ refers to any cell.

In eq. (9), the atomic-fragment energy $E_{(A,\mathbf{0})}$ is given by

$$E_{(A,\mathbf{0})} = \sum_{\mathbf{i},\mathbf{j}\in(A,\mathbf{0})} \sum_{\mathbf{ab}} E_{\mathbf{ij}}^{\mathbf{ab}}, \tag{10}$$

and involves an infinite sum over all virtual orbitals, which can be truncated in the same way as in the molecular DEC framework, giving[53,54]

$$E_{(A,\mathbf{0})} \simeq \sum_{\mathbf{i},\mathbf{j}\in(A,\mathbf{0})} \sum_{\mathbf{a},\mathbf{b}\in\mathcal{V}_{(A,\mathbf{0})}} E_{\mathbf{ij}}^{\mathbf{ab}}. \tag{11}$$

The finite set $\mathcal{V}_{(A,\mathbf{0})}$ is called the energy orbital space (EOS) and is composed of virtual orbitals assigned to centers in the vicinity of the atom $(A, \mathbf{0})$. The EOS is determined in a black-box manner by the fragment optimization, as explained in Section 2.4.

The second and third terms of eq. (9) correspond to the pair-fragment interaction energies, where the first atomic center $(A, \mathbf{0})$ is within the central unit cell, while the second one $(B, \mathbf{L}_B)$ runs over the entire system and is obtained by translation of the corresponding center in the central unit cell $(B, \mathbf{0})$. The pair-fragment interaction energy is approximated by

$$\Delta E_{(A,\mathbf{0}),(B,\mathbf{L}_B)} \simeq \left(\sum_{\mathbf{i}\in(A,\mathbf{0})} \sum_{\mathbf{j}\in(B,\mathbf{L}_B)} + \sum_{\mathbf{i}\in(B,\mathbf{L}_B)} \sum_{\mathbf{j}\in(A,\mathbf{0})} \right) \sum_{\substack{\mathbf{a},\mathbf{b}\in \\ \mathcal{V}_{(A,\mathbf{0})}\cup\mathcal{V}_{(B,\mathbf{L}_B)}}} E_{\mathbf{ij}}^{\mathbf{ab}}. \tag{12}$$

where EOSs obtained from atomic fragment optimizations are used. The lattice sum over $\mathbf{L}_B$ in eq. (9) is in principle infinite. However, the fast decay of the correlation energy with respect to the atomic center distance allows for truncation of this sum, as will be

detailed in Section 2.5. Note that the pair-fragment interaction energy in eq. (12) is obtained from a Boys-Bernardi counterpoise correction[68] for basis-set superposition-like errors: $\Delta E_{(A,\mathbf{0}),(B,\mathbf{L}_B)} = E_{(A,\mathbf{0}),(B,\mathbf{L}_B)} - E_{(A,\mathbf{0})} - E_{(B,\mathbf{L}_B)}$ where the atomic fragment energies are computed in the orbital spaces of the pair fragment. This procedure was found to be superior to the uncorrected approach for molecules[54].

The amplitudes for each atomic and pair fragment are obtained independently by solving the noncanonical MP2 equations,

$$0 = (\mathbf{ia}|\mathbf{jb}) + \sum_{\mathbf{c}} \left\{ t_{\mathbf{ij}}^{\mathbf{cb}} \langle \mathbf{a}|f|\mathbf{c}\rangle + t_{\mathbf{ij}}^{\mathbf{ac}} \langle \mathbf{c}|f|\mathbf{b}\rangle \right\} - \sum_{\mathbf{k}} \left\{ t_{\mathbf{ik}}^{\mathbf{ab}} \langle \mathbf{i}|f|\mathbf{k}\rangle + t_{\mathbf{kj}}^{\mathbf{ab}} \langle \mathbf{k}|f|\mathbf{j}\rangle \right\}, \tag{13}$$

where f is the periodic Fock operator. These equations are solved in an amplitude orbital space (AOS)[54] which includes, in addition to the EOS, both occupied and virtual orbitals centered on atoms in the vicinity of the EOS. Like the EOS, the AOS is determined independently for each atomic fragment during the fragment optimization procedure. For pair fragments, the union of the two AOSs is employed. The amplitude equations are solved noniteratively using a pseudocanonical formulation obtained by diagonalizing the occupied and virtual blocks of the Fock matrix within the AOS, as explained by Kristensen *et al.*[54]

2.3 Orbital assignment and orbital extents

The Wannier functions are expanded in contracted Gaussian-type AOs centered on atoms in unit cells within a domain $\mathcal{D}_{\mathbf{0}}$, which has the reference cell $\mathbf{0}$ at its center. Thus,

$$|W_{p\mathbf{0}}\rangle = \sum_{\mathbf{L}\in\mathcal{D}_{\mathbf{0}}} \sum_{\mu} |\chi_{\mu\mathbf{L}}\rangle C_{\mu p}^{-\mathbf{L}}, \tag{14}$$

where $C_{\mu p}^{-\mathbf{L}}$ is the Wannier coefficient associated with the AO $|\chi_{\mu\mathbf{L}}\rangle$ in cell $\mathbf{L}$. The size of the domain $\mathcal{D}_{\mathbf{0}}$ (Born-von Karman supercell) is determined by the k-point sampling used for generating the reference HF determinant in Bloch basis. Even for small unit cells, this often

means that each Wannier function is expanded in thousands of basis functions. However, the localized Wannier functions have non-negligible contributions only in a small subdomain of $\mathcal{D}_{\mathbf{0}}$ and a sufficently accurate description of the orbitals can be obtained by using a significantly smaller set of basis functions.

In the work of Jørgensen and co-workers, all molecular orbitals related to an atomic fragment are expanded on a common atomic fragment extent, which is defined as the union of all atomic extents in the AOS of that fragment[54]. Using a common atomic fragment extent for an entire atomic fragment would break the translation property (2) of the Wannier functions. Therefore, we expand all Wannier functions associated with the same atom on a single atomic extent. Orbital extents are computed as described in ref. 54, while atomic extents are set up as sketched in Algorithm 2. Instead of starting the optimization of each orbital extent from an empty set, as in ref. 54, a temporary atomic extent is used as a starting point (see Algorithm 2). This may give a different, and typically smaller, atomic extent compared with the approach in ref. 54. Before setting up the atomic extent, the Wannier functions associated with a considered atom are sorted with respect to decreasing orbital spread. After having determined an atomic extent, the coefficients of all Wannier functions associated with a given atom are fitted on the common atomic extent using least-squares optimization. The long-range tails of the Wannier functions, which are caused by the intercell orthogonality requirement (3), may be relatively poorly described on the atomic extent, depending on the chosen extent threshold. Since the correlation energy is mainly determined by the bulk of the orbitals, however, we do not expect this to pose significant problems unless very high accuracy is requested.

As in ref. 54, new atoms for the orbital extents are chosen from a list containing all atoms within $\mathcal{D}_{\mathbf{0}}$, sorted with respect to the absolute-value Mulliken projection. The Mulliken projection of orbital p in the reference cell on atom $\mathbf{A} \equiv (A, \mathbf{L}_A)$ in cell $\mathbf{L}_A$ is computed as

$$Q_p^{A,\mathbf{L}_A} = \sum_{\mu \in A} C_{\mu p}^{-\mathbf{L}_\mathbf{A}} \langle \chi_{\mu \mathbf{L}_\mathbf{A}} | W_{p\mathbf{0}} \rangle. \tag{15}$$

Each Wannier function is assigned to the atom with the greatest absolute-value Mulliken projection. If the greatest Mulliken projection of a given Wannier function is nearly the same on more than one atom, then the function is associated with the atom with the fewest previously assigned Wannier functions. Here, a relative difference smaller than 10^{-6} is used as the definition of 'nearly the same'.

Algorithm 2 Setup of an atomic extent $\mathcal{A}$ for an atom A, as done in the present study. The algorithm is otherwise as in ref. 54, but here the optimization starts from an initial extent, similarly as in the Boughton-Pulay procedure[69,70].

1: $\mathcal{A} \leftarrow$ An extent such that the sum of Mulliken projections for each Wannier function associated with atom A is at least 0.9.
2: Sort the Wannier functions according to decreasing spread.
3: **for** each Wannier function p associated with atom A **do**
4: Compute the orbital extent $\mathcal{O}(p)$ of orbital p, using $\mathcal{A}$ as an initial extent.
5: $\mathcal{A} \leftarrow \mathcal{A} \cup \mathcal{O}(p)$
6: **end for**

2.4 Atomic Fragment Optimization

For each atomic fragment, the EOS and AOS are optimized in an iterative way, using the same procedure as in the molecular DEC algorithm[54], that is, first expanding the EOS, and then the AOS. These successive expansions define a macro iteration and the fragment optimization is complete when the energy difference between two macro iterations is smaller than the FOT.

Before the atomic fragment optimization starts, every atom in the reference unit cell is associated with a list of neighbor atoms, sorted with respect to the distance from the considered central atom. In each step of the orbital-space expansion, the orbital space is increased with a constant volume until at least N_o occupied or virtual orbitals have been added.

2.5 Pair-fragment interaction energies

Because the MP2 pair-fragment interaction energy contributions decay asymptotically as r^{-6}, one can obtain accurate total pair interaction energies for molecular systems by including only pairs within a chosen cutoff radius[53,54]. For the same reason, for periodic systems, we need to compute pair-fragment interaction energy contributions explicitly within the MP2 approximation only for a finite number of pairs. Atoms in periodic systems occur regularly on a lattice and, for 3D systems, the number of atoms increases as r_{cut}^3 when the cutoff radius r_{cut} is increased. It is therefore both possible and necessary to interpolate pair correlation energies at large distances. In the present study, we compute pair energies only within a fixed cutoff distance, and in Sec. 4.3 we propose and test an algorithm to determine the cutoff distance in a black-box manner and consistent with the FOT. In the implementation of CRYSCOR[46], orbital-pair energies at large distances may be approximated by using extrapolations fitted to the London dispersion form Cr^{-6}, where r is the orbital-pair distance. Neese and co-workers have used collinear dipole-dipole interaction energy estimates to automatically determine which pairs are computed explicitly and which pairs are approximated by dipole-dipole interaction[15,71].

3 Computational details

The algorithm described above has been implemented in a stand-alone library called X-DEC. Orthonormal occupied and virtual Wannier functions, expanded in Gaussian-type AOs according to eq. (14), and the direct-space Fock matrix in AO basis must be provided as input. The X-DEC library is written in Python with C++ modules for efficient calculation of the required (**ia**|**jb**) integrals, which is the most time-consuming step of the algorithm. The two-electron integrals in AO basis are computed with Cauchy-Schwarz prescreening[72] using the Libint[73] library, and OpenMP[74] threading is employed. The fragment calculations are distributed to different compute nodes with the ZeroMQ[75] library, which offers the opportunity

to dynamically adjust the number of compute processes in a fault-tolerant manner. This is an important aspect of our hybrid parallelization strategy, since the required computational resources of an X-DEC calculation may not be easily estimated beforehand.

Since our main goals are validation of the X-DEC implementation and investigation of computational bottlenecks, we use the small split-valence 6-31G basis set[76] obtained from the Basis Set Exchange portal[77–79] for all calculations. For the same reasons, we choose simple periodic model systems for our tests: a 1D chain of Ne atoms, 2D and 3D collections of chains of Ne atoms, a 1D chain of ethylene molecules, and a 2D collection of ethylene chains.

The 1D Ne unit cell is defined to contain a single Ne atom with a lattice constant of 4.7 Bohr. The 2D Ne unit cell is quadratic and also contains a single Ne atom with lattice constant 4.7 Bohr. The 3D Ne unit cell is rectangular, again containing a single Ne atom, with lattice constants 4.7, 4.8, and 4.9 Bohr. The lattice constants are much smaller than the Ne-Ne distance in the face-centered cubic structure of solid neon and were chosen to increase the correlation energy. The 1D and 2D ethylene unit cells contain a single C_{2v} ethylene molecule with $R_{CC} = 2.5$ Bohr, $R_{CH} = 2.0$ Bohr, and $\theta_{HCH} = 120°$. The 1D system has a lattice constant of 7.0 Bohr and the periodic direction is chosen parallel to the carbon-carbon bond. The 2D system is obtained from the 1D system by repetition in the direction perpendicular to the plane of the ethylene molecule with lattice constant 10.0 Bohr. The lattice constant of 7.0 Bohr in the parallel direction implies that the nearest distance between carbon atoms in neighboring molecules is 4.5 Bohr, significantly less than the 7.2 Bohr in crystalline ethylene[80] and thus artificially increasing the correlation energy. The nearest carbon-carbon distance of 10.0 Bohr in the perpendicular direction, on other hand, is significantly greater—also compared with the equilibrium distance of 8.5 Bohr of the corresponding molecular dimer[81]. The 2D ethylene structure thus is expected to lead to more spread in the pair-fragment interaction energies, posing a challenge to accurate determination of a pair cutoff distance.

If a value is not explicitly given, the minimum number of added orbitals per expansion step in the atomic-fragment optimization, N_o, has the default value of ten for the neon and 1D ethylene systems and 15 for 2D ethylene. For comparison, Jørgensen and co-workers[61] chose an N_o parameter equal to 5η for molecular systems, with η the average number of molecular orbitals per atom. For the neon and ethylene systems, this corresponds to N_o values of 45 and 33, respectively. Our fragment optimizations thus are more fine-grained than the molecular ones, which potentially leads to smaller atomic fragments but also increases the risk of adding less important orbitals before the more important ones, giving false convergence.

The HF calculations were performed with the CRYSTAL[65] program, using a Brillouin-zone shrinking factor `SHRINK` of 8 (i.e., $8\times8\times8$ k-point sampling in 3D), a convergence threshold on the energy of 10^{-12} Ha, and an integral threshold of 10^{-16} Ha (all `ITOL` parameters set to 16). The wannierization was performed with the Brillouin-zone shrinking factor `NEWK` set to 16 (i.e., the converged Bloch orbitals were converted to a $16\times16\times16$ k-point grid before starting the localization-wannierization procedure). The Fock, overlap, and molecular coefficient matrices were obtained through a locally modified version of the CRYSTAL interface program `cryapi_inp`[82].

For validation purposes, we used the local MP2 implementation of the CRYSCOR program[48]. In order to compare the energies obtained with X-DEC for a given FOT, the parameters of CRYSCOR were pushed towards the convergence limit.

Calculations for clusters of 3, 5, 10, 15, 20, and 30 ethylene molecules were performed with the molecular DEC-MP2 implementation in the quantum chemistry package LSDalton[83,84]. In these computations, the FOT was set to 10^{-5} Ha, the HF convergence threshold (gradient norm) was set to 10^{-8} Ha, and the no-core approximation was switched off. Default values were used for all other parameters. The cluster correlation energies E_N were fitted to the linear expression $a\frac{1}{N}+b$, where N is the number of monomers. The extrapolated energy per monomer is equal to b, which corresponds to the limit when N approaches infinity.

4 Results

4.1 Validation of X-DEC-MP2 correlation energies

In Table 1, X-DEC-MP2 correlation energies for neon systems in one and two dimensions, and a one-dimensional periodic chain of ethylene molecules, are compared with results[85] from CRYSCOR[48]. For the 1D systems, extrapolated correlation energies obtained from finite-chain molecular DEC calculations with LSDALTON[83,84] are also shown. The results in Table 1 were obtained with an extent tolerance 10^{-7} and a pair cutoff at 30.0 Bohr.

For the neon systems presented in Table 1, the correlation energy obtained with X-DEC and CRYSCOR are within 1 FOT of one another. The difference between the X-DEC and extrapolated molecular results for the 1D systems is at most four times the FOT. Both with FOT 10^{-4} and 10^{-5}, the difference of the X-DEC-MP2 correlation energy for 1D ethylene was below 10^{-4} Ha. Despite the simplicity of the model systems used here, these results confirm that the X-DEC algorithm provides error control when computing the MP2 energy for periodic systems.

Table 1: Comparison of MP2 correlation energies in milli-Hartree, as obtained using X-DEC, CRYSCOR[48], and LSDalton[83,84]. The FOT is given in Hartree.

System	X-DEC	(FOT)	CRYSCOR	LSDalton
1D neon	−114.36319	(10^{-7})	−114.36316	−114.36285
2D neon	−114.470	(10^{-4})	−114.463	–
	−114.470	(10^{-5})		
1D ethylene	−187.191	(10^{-4})	−187.248	−187.303
	−187.344	(10^{-5})		

The key to the correlation-energy error control is the atomic fragment optimization. As can be seen from Table 2, the minimum number of orbitals, N_o, per expansion step in the atomic fragment optimization must be sufficiently large to avoid that the algorithm stops before the energy is truly converged to within one FOT. In the 1D ethylene system, the occupied and virtual Wannier orbitals are all assigned to one of the two carbon atoms,

which are atoms 4 and 5 in our input sequence; hence the carbon atoms are referred to by these numbers in the following. In an exact treatment (or using full space group symmetry), the two carbon atoms are exactly equivalent. Not explicitly using point-group symmetry, the X-DEC algorithm may treat the carbon atoms differently, leading to fragment energies that differ beyond the FOT. With a FOT of 10^{-5} Ha the optimization for atom 5 stops at -63.15 mHa when $N_o = 2$, while an atomic fragment energy of -63.24 mHa is obtained when at least ten orbitals is required per expansion step. Thus, in the former case, the atomic fragment energy is not truly converged to within one FOT. This shows that sorting of orbitals based on distance is not necessarily the optimal choice, as sometimes orbitals farther from the center may give a larger contribution to the energy. Kristensen and co-workers[61] handled this problem by first using a large expansion parameter N_o and then reducing the orbital spaces with a binary search algorithm. As shown in Table 2, a minimum expansion parameter N_o of 20 gives the same atomic fragment energies to within one FOT as when the parameter is set to ten. We therefore assume that the atomic fragment energy for the 1D ethylene system is converged to within the given FOT value when minimum ten orbitals is required per expansion step.

Table 2: Atomic-fragment and total correlation energies of 1D ethylene, as obtained using different FOTs and minimum orbital expansion parameters N_o. The symbols E_4 and E_5 denote atomic fragment energies for the two (carbon) atomic fragments in the system, and $E_{\text{X-DEC-MP2}}$ is the total X-DEC-MP2 correlation energy. In these computations, the extent tolerance was 10^{-7} and the pair cutoff was set to 30.0 Bohr.

FOT/Ha	N_o	E_4/mHa	E_5/mHa	$E_{\text{X-DEC-MP2}}$/mHa
10^{-4}	2	-63.22	-63.09	-187.11
10^{-4}	10	-63.15	-63.22	-187.19
10^{-4}	20	-63.23	-63.24	-187.33
10^{-5}	2	-63.22	-63.15	-187.19
10^{-5}	10	-63.24	-63.24	-187.34
10^{-5}	20	-63.24	-63.24	-187.34

4.2 Analysis of atomic fragments

Figures 1 and 2 show the number of atoms in the different orbital spaces for the 1D and 2D neon systems for different FOT values. For too big a FOT, the orbital spaces only contain one atom, which means that the atomic-fragment MP2 correlation energy only takes into account the underlying periodicity via the shape of the Wannier functions. As the FOT decreases, the occupied AOS includes first the nearest neighbor atoms, then the second-nearest neighbors. For the virtual EOS and AOS, the number of included atoms corresponds to second- and third-nearest neighbors, respectively. The change in the correlation energy is significant, ca. 0.5 mHa, when the nearest neighbors are included, but is then minimal (10^{-8} Ha) when the second-nearest neighbors are added in the occupied AOS. Even larger virtual spaces were also explored for Neon 1D (fourth- and sixth-nearest neighbors for the EOS and AOS respectively) but this had no significant effect on the total correlation energy.

Converging the correlation energy below 1 mHa thus requires inclusion of at least second-nearest neighbors in the virtual AOS. While this does not pose a severe computational challenge in 1D, it quickly becomes a problem in 2D and 3D. Table 3 shows the AOS sizes (union of the two atomic fragment AOSs) and timing of selected pair fragments at similar distances in the 1D, 2D, and 3D neon systems. Evidently, not only the number of orbitals in the AOS but also the atomic fragment extents make the calculations expensive. Note that we use neither screening of the AO-to-Wannier transformation nor integral approximations, which would decrease the timing by 1–2 orders of magnitude but leave the increasing trend unchanged.

For 2D ethylene, all occupied Wannier orbitals are assigned to the two carbon atoms, while the virtual Wannier orbitals are distributed among all six atoms. This implies that only the two carbon atomic fragments are optimized, giving AOSs composed of occupied Wannier orbitals on 24 (carbon) atoms and virtual Wannier orbitals on 38 atoms with a FOT of 10^{-4} Ha. With an extent tolerance of 10^{-4}, these orbitals are expanded on basis functions on more than 170 atoms. Calculations of pair fragments thus involve systems with hundreds of

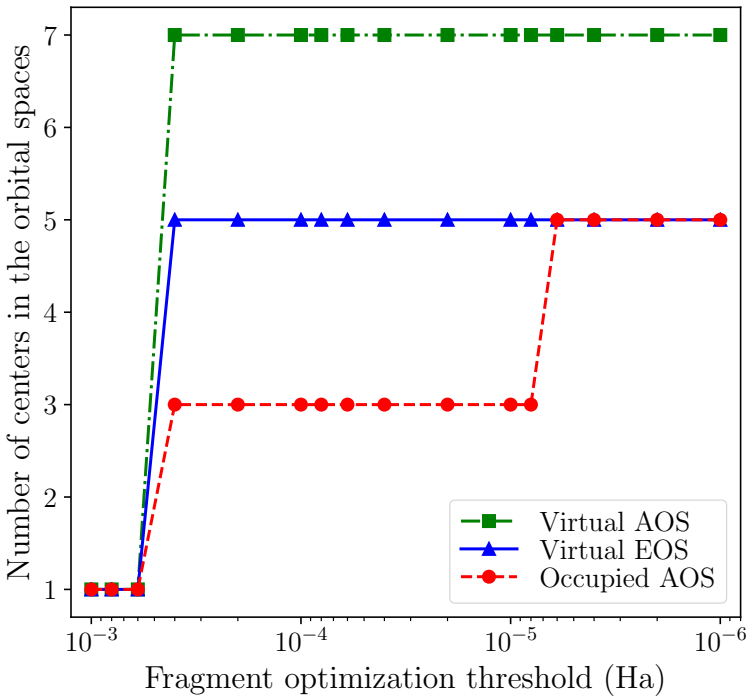

Figure 1: Number of atomic centers in the different orbital spaces as a function of the FOT, here for the 1D Ne system. The atomic-fragment MP2 correlation energies for the FOT values 10^{-3}, 10^{-4}, and 10^{-6} Ha are -0.1138401565, -0.1143028172, and -0.1143028173 Ha, respectively. These results were obtained with the minimum number of orbitals per expansion $N_o = 2$.

Table 3: Sizes of pair fragments for the 1D, 2D, and 3D Ne systems with FOT 10^{-4} Ha and extent tolerance 10^{-4}. The pair distance is denoted by r. N_{occ} is the number of occupied orbitals and N_{virt} the number of virtual orbitals of the AOS, while N_{basis} is the number of basis functions used to span the AOS.

System	r (Bohr)	N_{occ}	N_{virt}	N_{basis}	Time (s)
1D neon	23.5	30	72	180	5
2D neon	23.5	50	104	432	382
3D neon	24.0	30	152	1026	11212

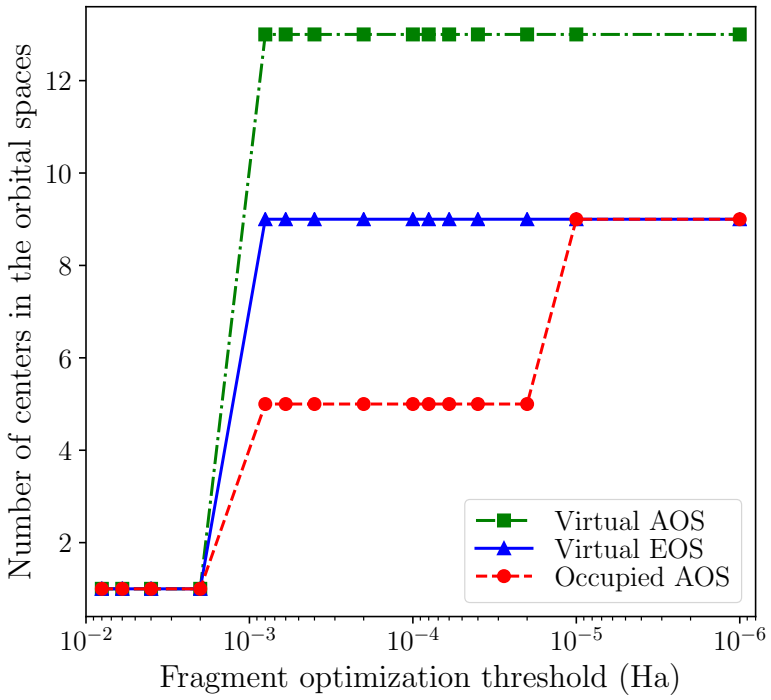

Figure 2: Number of atomic centers in the different orbital spaces as a function of the FOT, here for the 2D Ne system. The atomic-fragment MP2 correlation energies for the FOT values $2 \cdot 10^{-3}$, 10^{-4}, and 10^{-5} Ha are -0.113420366, -0.114344148, and -0.114344155 Ha, respectively.

occupied and virtual orbitals and more than 1000 basis functions, which is clearly too big for routine calculations. Again, integral approximations and efficient screening procedures would speed up the calculations tremendously but would not address the central problem: the AOS is too large.

While we are currently investigating different options for spanning the virtual space, which is the main problem of the AOS, we will in the following focus on the major computational bottleneck, both in terms of memory consumption and time, of the X-DEC algorithm, which is the pair-fragment calculations. Even with a reduced AOS, which is the subject of a forthcoming publication, the sheer number of pairs is a challenge. This motivates a study of pair interpolation procedures.

4.3 Determination of the pair cutoff distance

Provided the distributions $|\mathbf{ia})$ and $|\mathbf{jb})$ are sufficiently separated, it follows from the bipolar multipole expansion[86] of the integrals $(\mathbf{ia}|\mathbf{jb})$ that, to leading order, the pair-fragment interaction energies decay as[54,55] r^{-6}, where r is the pair distance. This implies that pairs beyond a certain cutoff distance r_{cut} may either be neglected or approximated by an expression of the form $C_{AB}r_{AB}^{-6}$, with the constant C_{AB} determined for each pair of atoms through simple least-squares fitting using just a few points. The challenge is to determine r_{cut} in a black-box manner consistent with the specified FOT.

We first verify that the pair-fragment interaction energies indeed decay as expected (see Figures 3 and 4 for 3D neon and 2D ethylene, respectively). For 3D neon, all pair energies lie very close to the function $0.2 \cdot r^{-6.2}$ obtained by least-squares fitting. The pair energies for the 2D ethylene system are more spread but still show an overall decay approximately given by $141.1 \cdot r^{-6.4}$.

With a given cutoff distance r_{cut}, the total pair-fragment interaction energy may be

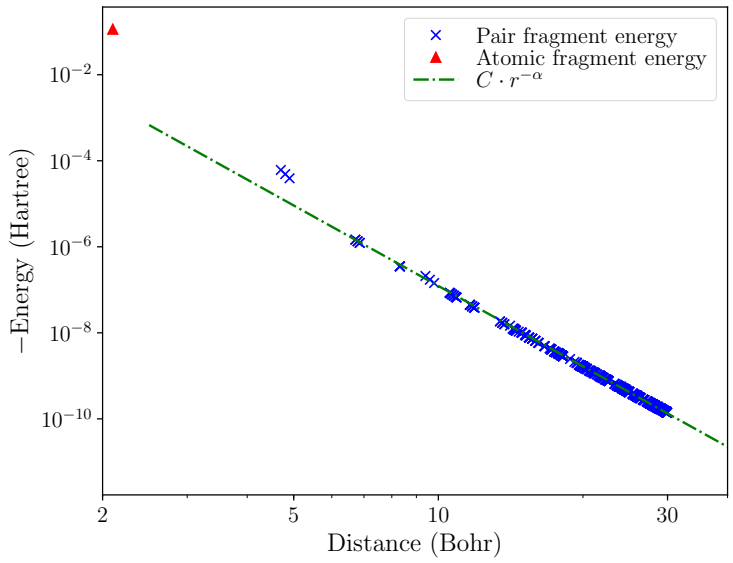

Figure 3: Atomic and pair-fragment correlation energies for 3D neon as a function of the pair distance. The dashed line represents a least-squares fitting to the formula $Cr^{-\alpha}$, which gives the parameters $C = 0.2006$ and $\alpha = 6.2126$. The X-DEC-MP2 calculation was done with FOT $= 10^{-4}$ Ha, extent tolerance 10^{-4}, and pair cutoff distance of 30.0 Bohr. Both axes are given in logarithmic scale.

written as

$$\Delta E_{\text{pairs}} = \sum_A \left(\sum_B \Delta E^{<}_{AB} + \sum_B \Delta E^{>}_{AB} \right), \tag{16}$$

where the superscripts $<$ and $>$ denote contributions from pairs within and beyond the cutoff distance, respectively, and

$$\Delta E^{<}_{AB} = \Delta E_{(A,\mathbf{0}),(B,\mathbf{0})} + \sum_{\mathbf{L_B}\neq\mathbf{0}}^{r_{(A,\mathbf{0}),(B,\mathbf{L}_B)}<r_{\text{cut}}} \Delta E_{(A,\mathbf{0}),(B,\mathbf{L}_B)}, \tag{17}$$

$$\Delta E^{>}_{AB} = \sum_{\mathbf{L_B}}^{r_{(A,\mathbf{0}),(B,\mathbf{L}_B)}\geq r_{\text{cut}}} \Delta E_{(A,\mathbf{0}),(B,\mathbf{L}_B)}. \tag{18}$$

Note that that the first term of eq. (17) is included only if $B > A$, in agreement with eq. (9).

In the molecular DEC-MP2 algorithm, $\Delta E^{>}_{AB}$ is neglected for all A, B. For an extended system, however, this may not be a sufficiently good approximation. Introducing the long-

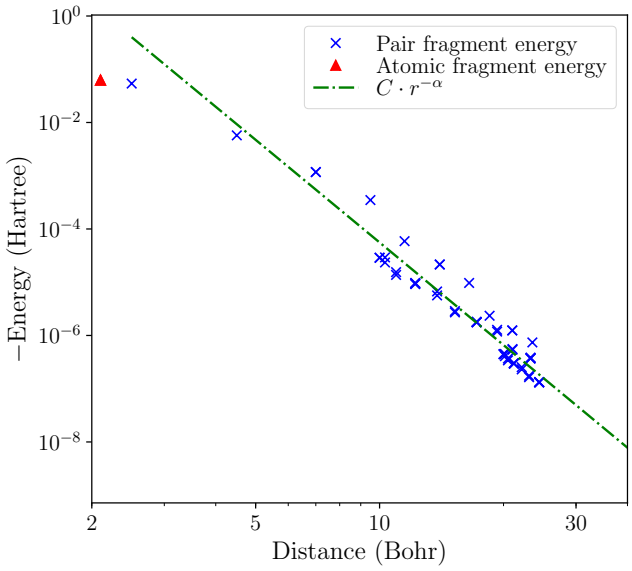

Figure 4: Atomic and pair-fragment correlation energies for 2D ethylene as a function of the pair distance. The dashed line represents a least-squares fitting to the formula $Cr^{-\alpha}$, which gives the parameters $C = 141.0948$ and $\alpha = 6.4026$. The calculation was done with FOT $= 10^{-4}$ Ha, extent tolerance 10^{-4}, and a pair cutoff distance of 25.0 Bohr. Both axes are given in logarithmic scale.

range approximation

$$\Delta E^{>}_{AB} \simeq C_{AB} \sum_{\mathbf{L_B}}^{r_{(A,\mathbf{0}),(B,\mathbf{L}_B)} \geq r_{\text{cut}}} r^{-6}_{(A,\mathbf{0}),(B,\mathbf{L}_B)} \tag{19}$$

and replacing the summation by integration[46], we get

$$\Delta E^{>}_{AB} \approx \begin{cases} \frac{1}{5} C_{AB} r^{-5}_{\text{cut}} & \text{(1D)} \\ \frac{\pi}{2} C_{AB} r^{-4}_{\text{cut}} & \text{(2D)} \\ \frac{4\pi}{3} C_{AB} r^{-3}_{\text{cut}} & \text{(3D)}, \end{cases} \tag{20}$$

which may not be negligible relative to the FOT unless a very large cutoff distance is chosen. On the other hand, computational performance considerations force us to choose r_{cut} as small as possible.

We therefore investigate an algorithm to identify a reasonable cutoff distance in a predefined interval $[r_1, r_2]$. Spline interpolation, and possibly also least-squares fitting using

eqs. (19) and (20), may be used to estimate the energy contributions outside the obtained cutoff. First, a limited number of pairs with distances between r_1 and r_2 are computed. A cubic spline is then obtained using the computed energies, and, depending on the radius r_2, possibly also using one or a few energies at pair distances greater than r_2. The extra data points at large pair distances may be needed to avoid unphysical oscillations in the interpolation around r_2. Second, a few pairs beyond r_2 are computed explicitly and the constants C_{AB} are determined by least-squares fitting for each pair A, B. The total pair energy beyond r_2, $E_{AB}^{r_2}$, is estimated using eq. (20) with r_2 in place of r_{cut}. To find a pair cutoff r_{cut}, pair energies from the cubic spline interpolation, ordered with respect to decreasing distance, are added until the sum of $E_{AB}^{r_2}$ and the interpolated energies is approximately equal to a tolerance, which should be determined indirectly by the FOT. The distance at which this happens defines the cutoff distance. Finally, we can evaluate the total correlation energy from eq. (9) via explicit calculation of interaction energies for all pairs within the cutoff distance.

Using cubic spline interpolation, this algorithm does not rely on an assumed (polynomial) decay behavior in the crucial interval $[r_1, r_2]$, which implies that a rather large interval may be used without a severe computational penalty. In the following, we test this algorithm on the 3D Ne and 2D ethylene model systems.

For 2D ethylene, the interpolation limits r_1 and r_2 were chosen to be 0.0 and 25.0 Bohr, respectively. In this interval, the 2D ethylene system has 15 pairs with the carbon atoms $A = 4$ and $B = 5$. Out of these, 6 pairs with a pair distance smaller than 25.0 Bohr, and additionally one pair with a distance of approximately 41.6 Bohr, were used for the cubic splines interpolation. The splines interpolation was computed with the Scipy library `interpolate.splrep`[87] and a smoothing factor[88] of 1.0. The used spline data points and the resulting spline interpolation for 2D ethylene are shown in Figure 5. To test the quality of the spline interpolation, a least-squares fit of a function $Cr^{-\alpha}$ was obtained based on interpolation points between 25.0 and 40.0 Bohr. This gave $\alpha \approx 6.7$, indicating that the

spline interpolation is not very far from a r^{-6} decay in the considered interval.

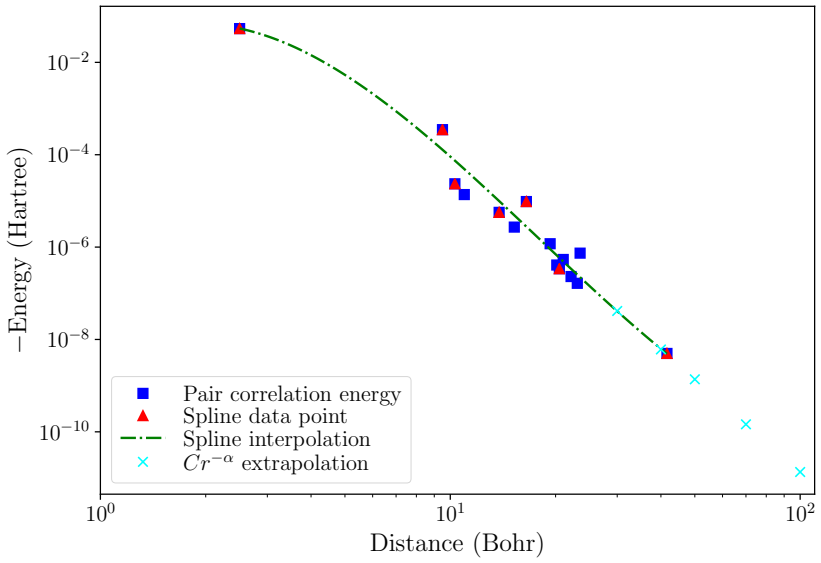

Figure 5: Illustration of pair interpolation for the carbon atoms $A = 4$ and $B = 5$ of 2D ethylene. $C = -292.929$ Ha and $\alpha = 6.668$.

Starting from this spline interpolation of 2D ethylene, pair cutoff distances were determined for given tolerances. Pair contributions beyond 25.0 Bohr were neglected. As explained above, interpolated energies for all pairs within $[r_1, r_2]$ were sorted with respect to the corresponding pair distance. When summing energies according to decreasing distance, the cutoff r_{cut} was determined as the pair distance of the first pair giving a sum larger than the tolerance. The total pair energy for the pairs (A, B), $A = 4$ and $B = 5$, of the 2D ethylene system was then approximated as the sum of all "exact" X-DEC-MP2 energies within r_{cut} plus the sum of interpolated energies for all pairs within the interval $[r_{\text{cut}}, r_2]$, where r_2 is as defined above. Letting E_{AB}^{approx} be the approximated energy and E_{AB}^{exact} the sum of all "exact" X-DEC-MP2 pair energies within $[r_1, r_2]$, the error of the approximation using interpolated values within $[r_{\text{cut}}, r_2]$ is

$$\varepsilon = |E_{AB}^{\text{approx}} - E_{AB}^{\text{exact}}|. \tag{21}$$

As shown in Table 4, for tolerances 10^{-4}, 10^{-5}, and 10^{-6} Ha, the obtained errors are all smaller than the tolerance values, and the pair cutoff r_{cut} is found to decrease significantly with decreasing tolerance. The obtained pair cutoffs are comparable to orbital-pair cutoffs used in typical CRYSCOR calculations[46], but with our approach the pair cutoffs are determined based on a given error threshold. As seen from Table 5, for 3D neon with r_1 and r_2 set to 0.0 and 30.0 Bohr, respectively, the errors for this system are also smaller than the given tolerances.

These results indicate that the proposed algorithm may be useful for computing the pair correlation energy to a precision restricted by the FOT, without having to use a too large pair cutoff. By reducing the pair cutoff, the number of pairs for which the energy has to be computed explicitly may be reduced significantly. This would give a reduction in computing time roughly proportional to the reduction of number of pairs. The relative reduction in computing time obtained with a smaller pair cutoff is generally larger for three-dimensional systems, because the number of pairs increases as r_{cut}^d, where r_{cut} is the cutoff radius and d the number of periodic dimensions. This is here illustrated by the speedups for 2D ethylene and 3D neon, given in Tables 4 and 5. For the 3D neon system, the number of needed spline data points is smaller compared to the total number of pairs in a given interval, and therefore the speedup is more significant than for 2D ethylene.

Table 4: Pair approximation based on a given tolerance parameter for 2D ethylene.

Tolerance (Ha)	Cutoff (Bohr)	Error (Ha)	Speedup
10^{-4}	10.3	$3.5 \cdot 10^{-5}$	3.0
10^{-5}	15.2	$7.1 \cdot 10^{-6}$	2.3
10^{-6}	21.0	$3.2 \cdot 10^{-7}$	1.5

For the spline interpolations, data points are chosen so that points with the shortest distances are close to each other, and the interval between the chosen distances increase with increasing distance. The reason for choosing points in this way, is that the pair data in log-log plots generally have a larger curvature for smaller pair distances, and at large pair

Table 5: Pair approximation based on a given tolerance parameter for 3D neon.

Tolerance (Ha)	Cutoff (Bohr)	Error (Ha)	Speedup
10^{-4}	4.7	$6.4 \cdot 10^{-5}$	116
10^{-5}	6.7	$8.2 \cdot 10^{-8}$	91
10^{-6}	12	$5.5 \cdot 10^{-9}$	22

distances the curve approaches linear, according to the multipole expansion approximation[55]. As mentioned above, it is also important to include data points with large pair distances. When choosing points following these principles, points may still be picked in different ways, and therefore give different results. In the 2D ethylene example, we always used the three first and the last pairs. With those points fixed, choosing the three remaining points in several different ways always gave an error smaller than the tolerance. For the tolerance 10^{-4}, the errors changed in the interval $2.4 - 3.6 \cdot 10^{-5}$, for the tolerance 10^{-5} the errors were in the range $6.3 - 9.4 \cdot 10^{-6}$, and spline interpolations with a tolerance 10^{-6} gave errors between $8.9 \cdot 10^{-8}$ and $7.7 \cdot 10^{-7}$. Similarly, for 3D neon, the errors were always smaller than the tolerance when fixing the first and last spline data point and choosing the four intermediate points in different ways. Thus, as long as the general principles described above were followed, the interpolation seemed to be stable with respect to exactly which points were used.

5 Conclusions

We have described a generalization of the molecular DEC-CC approach to extended systems with periodic boundary conditions. The X-DEC algorithm follows closely the molecular DEC algorithm and we have shown that the crucial aspect of correlation energy error control is conserved. Although we have only presented results at the MP2 level of theory, the X-DEC framework can be easily applied to other CC models such as the random-phase approximation, CCSD, and CCSD(T). The X-DEC algorithm is currently too computation-

ally demanding, however, and we have therefore used very simple model systems and a small split-valence basis set only. The main challenges are the AOS sizes and the associated atomic extents used to expand the AOS orbitals, and the sheer number of pair fragments, which generally is much greater for extended systems than for molecules. We have proposed and investigated a black-box approach based on cubic splines interpolation to determine a pair cutoff distance with a low computational cost, while still maintaining the error control characteristic of the DEC approach.

The AOS problem is more subtle and has not been solved in the present work. There are two aspects of the AOS that must be considered. First, *all* the occupied Wannier functions must be as local as possible, allowing the AOS to contain only virtual Wannier functions in a close vicinity of the occupied ones. If just one occupied Wannier function is rather poorly localized, the virtual part of the AOS may quickly become intractable. Second, the virtual Wannier functions must be sufficiently localized to allow a physically reasonable assignment to an atom. At the same time, however, the virtual Wannier functions of the reference cell must remain orthogonal to the total occupied space *and* span as much of the total virtual space as possible in order to get as compact a representation of the correlation energy as possible. While recent advances in orbital localization, both for molecules[62] and periodic systems[89], may be leveraged for the occupied Wannier functions, the conditions on the virtual Wannier functions seem difficult to fulfill using existing orbital localization criteria. We are currently investigating different options and will report on our findings in the near future.

Even with more compact AOSs, the X-DEC-MP2 algorithm presented here can be easily accelerated using integral approximations such as the resolution-of-the-identity (RI) approximation.[90] This has recently been successfully done for molecular DEC-MP2 calculations[57] and our own unpublished pilot implementation promises speedups of 1–2 orders of magnitude.

In conclusion, the X-DEC algorithm provides a framework for periodic local CC calculations with correlation energy error control but more work is required to accelerate the

algorithm.

Acknowledgement

This work was partially supported by the Research Council of Norway through its Centres of Excellence scheme, project number 262695, and through the Research Grant No. 240698. This work has received support from the Norwegian Supercomputing Program (NOTUR) through a grant of computer time (Grant No. NN4654K). The authors would like to thank L. Maschio for many helpful discussions and help running CRYSTAL and CRYSCOR. I.-M. Høyvik is thanked for help with orbital localization in LSDalton and E. Valeev for help with Libint. A. Scemama and B. Barbot are gratefully acknowledged for help with ZeroMQ. Finally, G.B. thanks L. Wirz for discussions related to computing.

References

(1) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291–352.

(2) Shavitt, I.; Bartlett, R. J. *Many-Body Methods in Chemistry and Physics*; Cambridge University Press, 2009.

(3) Rohwedder, T.; Schneider, R. Error estimates for the Coupled Cluster method. *ESAIM: Math. Model. Numer. Anal.* **2013**, *47*, 1553–1582.

(4) Sinanoğlu, O. Many-Electron Theory of Atoms, Molecules and Their Interactions. *Adv. Chem. Phys.* **1964**, *6*, 315–412.

(5) Nesbet, R. K. Electron Correlation in Atoms and Molecules. *Adv. Chem. Phys.* **1965**, *9*, 321–363.

(6) Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151–154.

(7) Sæbø, S.; Pulay, P. Local configuration interaction: An efficient approach for larger molecules. *Chem. Phys. Lett.* **1985**, *113*, 13–18.

(8) Sæbø, S.; Pulay, P. Fourth-order Møller–Plessett perturbation theory in the local correlation treatment. I. Method. *J. Chem. Phys.* **1987**, *86*, 914–922.

(9) Sæbø, S.; Pulay, P. The local correlation treatment. II. Implementation and tests. *J. Chem. Phys.* **1988**, *88*, 1884–1890.

(10) Sæbø, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213–236.

(11) Korona, T.; Kats, D.; Schütz, M.; Adler, T. B.; Liu, Y.; Werner, H.-J. In *Linear-Scaling Techniques in Computational Chemistry and Physics*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Challenges and Advances in Computational Chemistry and Physics; Springer Netherlands: Dordrecht, 2011; Vol. 13; Chapter 14, pp 345–407.

(12) Werner, H.-J.; Knizia, G.; Krause, C.; Schwilk, M.; Dornbach, M. Scalable Electron Correlation Methods I.: PNO-LMP2 with Linear Scaling in the Molecular Size and Near-Inverse-Linear Scaling in the Number of Processors. *J. Chem. Theory Comput.* **2015**, *11*, 484–507.

(13) Ma, Q.; Werner, H.-J. Scalable Electron Correlation Methods. 2. Parallel PNO-LMP2-F12 with Near Linear Scaling in the Molecular Size. *J. Chem. Theory Comput.* **2015**, *11*, 5291–5304.

(14) Schwilk, M.; Ma, Q.; Köppl, C.; Werner, H.-J. Scalable Electron Correlation Methods.

3. Efficient and Accurate Parallel Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD). *J. Chem. Theory Comput.* **2017**, *13*, 3650–3675.

(15) Pinski, P.; Riplinger, C.; Valeev, E. F.; Neese, F. Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals. *J. Chem. Phys.* **2015**, *143*, 034108.

(16) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **2016**, *144*, 024109.

(17) Pavošević, F.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. SparseMaps—A systematic infrastructure for reduced-scaling electronic structure methods. IV. Linear-scaling second-order explicitly correlated energy with pair natural orbitals. *J. Chem. Phys.* **2016**, *144*, 144109.

(18) Pavošević, F.; Peng, C.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. SparseMaps—A systematic infrastructure for reduced scaling electronic structure methods. V. Linear scaling explicitly correlated coupled-cluster method with pair natural orbitals. *J. Chem. Phys.* **2017**, *146*, 174108.

(19) Hirata, S.; Grabowski, I.; Tobita, M.; Bartlett, R. J. Highly accurate treatment of electron correlation in polymers: coupled-cluster and many-body perturbation theories. *Chem. Phys. Lett.* **2001**, *345*, 475–480.

(20) Hirata, S.; Podeszwa, R.; Tobita, M.; Bartlett, R. J. Coupled-cluster singles and doubles for extended systems. *J. Chem. Phys.* **2004**, *120*, 2581–2592.

(21) Keçeli, M.; Hirata, S. Fast coupled-cluster singles and doubles for extended systems:

Application to the anharmonic vibrational frequencies of polyethylene in the Γ approximation. *Phys. Rev. B* **2010**, *82*, 115107.

(22) Ohnishi, Y.-Y.; Hirata, S. Hybrid coupled-cluster and perturbation method for extended systems of one-dimensional periodicity. *J. Chem. Phys.* **2011**, *135*, 094108.

(23) Marsman, M.; Grüneis, A.; Paier, J.; Kresse, G. Second-order Møller–Plesset perturbation theory applied to extended systems. I. Within the projector-augmented-wave formalism using a plane wave basis set. *J. Chem. Phys.* **2009**, *130*, 184103.

(24) Grüneis, A.; Marsman, M.; Kresse, G. Second-order Møller-Plesset perturbation theory applied to extended systems. II. Structural and energetic properties. *J. Chem. Phys.* **2010**, *133*, 074107.

(25) Grüneis, A.; Booth, G. H.; Marsman, M.; Spencer, J.; Alavi, A.; Kresse, G. Natural Orbitals for Wave Function Based Correlated Calculations Using a Plane Wave Basis Set. *J. Chem. Theory Comput.* **2011**, *7*, 2780–2785.

(26) Booth, G. H.; Grüneis, A.; Kresse, G.; Alavi, A. Towards an exact description of electronic wavefunctions in real solids. *Nature* **2013**, *493*, 365–370.

(27) McClain, J.; Sun, Q.; Chan, G. K.-L.; Berkelbach, T. C. Gaussian-Based Coupled-Cluster Theory for the Ground-State and Band Structure of Solids. *J. Chem. Theory Comput.* **2017**, *13*, 1209–1218.

(28) Stanton, J. F.; Bartlett, R. J. The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *J. Chem. Phys.* **1993**, *98*, 7029–7039.

(29) Krylov, A. I. Equation-of-Motion Coupled-Cluster Methods for Open-Shell and Electronically Excited Species: The Hitchhiker's Guide to Fock Space. *Annu. Rev. Phys. Chem.* **2008**, *59*, 433–462.

(30) Ayala, P. Y.; Kudin, K. N.; Scuseria, G. E. Atomic orbital Laplace-transformed second-order Møller–Plesset theory for periodic systems. *J. Chem. Phys.* **2001**, *115*, 9698–9707.

(31) Almlöf, J. Elimination of energy denominators in Møller—Plesset perturbation theory by a Laplace transform approach. *Chem. Phys. Lett.* **1991**, *181*, 319–320.

(32) Häser, M.; Almlöf, J. Laplace transform techniques in Møller-Plesset perturbation theory. *J. Chem. Phys.* **1992**, *96*, 489–494.

(33) Häser, M. Møller-Plesset (MP2) perturbation theory for large molecules. *Theor. Chim. Acta* **1993**, *87*, 147–173.

(34) Del Ben, M.; Hutter, J.; VandeVondele, J. Second-Order Møller-Plesset Theory in the Condensed Phase: An Efficient and Massively Parallel Gaussian and Plane Waves Approach. *J. Chem. Theory Comput.* **2012**, *8*, 4177–4188.

(35) Del Ben, M.; Hutter, J.; VandeVondele, J. Forces and stress in second order Møller-Plesset perturbation theory for condensed phase systems within the resolution-of-identity Gaussian and plane waves approach. *J. Chem. Phys.* **2015**, *143*, 102803.

(36) Kotliar, G.; Savrasov, S. Y.; Haule, K.; Oudovenko, V. S.; Parcollet, O.; Marianetti, C. A. Electronic structure calculations with dynamical mean-field theory. *Rev. Mod. Phys.* **2006**, *78*, 865–951.

(37) Libisch, F.; Huang, C.; Carter, E. A. Embedded correlated wavefunction schemes: theory and applications. *Acc. Chem. Res.* **2014**, *47*, 2768–2775.

(38) Sun, Q.; Chan, G. K.-L. Quantum Embedding Theories. *Acc. Chem. Res.* **2016**, *49*, 2705–2712.

(39) Stoll, H. Correlation energy of diamond. *Phys. Rev. B* **1992**, *46*, 6700–6704.

(40) Stoll, H.; Paulus, B.; Fulde, P. On the accuracy of correlation-energy expansions in terms of local increments. *J. Chem. Phys.* **2005**, *123*, 144108.

(41) Paulus, B. The method of increments—a wavefunction-based ab initio correlation method for solids. *Phys. Rep.* **2006**, *428*, 1–52.

(42) Nolan, S. J.; Gillan, M. J.; Alfè, D.; Allan, N. L.; Manby, F. R. Calculation of properties of crystalline lithium hydride using correlated wave function theory. *Phys. Rev. B* **2009**, *80*, 165109.

(43) Collins, M. A. Ab initio lattice dynamics of nonconducting crystals by systematic fragmentation. *J. Chem. Phys.* **2011**, *134*, 164110.

(44) Pisani, C.; Busso, M.; Capecchi, G.; Casassa, S.; Dovesi, R.; Maschio, L.; Zicovich-Wilson, C.; Schütz, M. Local-MP2 electron correlation method for nonconducting crystals. *J. Chem. Phys.* **2005**, *122*, 094113.

(45) Maschio, L.; Usvyat, D.; Manby, F. R.; Casassa, S.; Pisani, C.; Schütz, M. Fast local-MP2 method with density-fitting for crystals. I. Theory and algorithms. *Phys. Rev. B* **2007**, *76*, 075101.

(46) Pisani, C.; Maschio, L.; Casassa, S.; Halo, M.; Schütz, M.; Usvyat, D. Periodic local MP2 method for the study of electronic correlation in crystals: Theory and preliminary applications. *J. Comput. Chem.* **2008**, *29*, 2113–2124.

(47) Halo, M.; Casassa, S.; Maschio, L.; Pisani, C.; Amor, N. B.; Borini, S.; Maynau, D.; Rossi, E. Periodic local-MP2 computational study of crystalline neon. *Phys. Chem. Chem. Phys.* **2009**, *11*, 586–592.

(48) Pisani, C.; Schütz, M.; Casassa, S.; Usvyat, D.; Maschio, L.; Lorenz, M.; Erba, A. Cryscor: a program for the post-Hartree-Fock treatment of periodic systems. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7615–7628.

(49) Usvyat, D.; Maschio, L.; Schütz, M. Periodic local MP2 method employing orbital specific virtuals. *J. Chem. Phys.* **2015**, *143*, 102805.

(50) Wannier, G. H. Dynamics of Band Electrons in Electric and Magnetic Fields. *Rev. Mod. Phys.* **1962**, *34*, 645–655.

(51) Marzari, N.; Mostofi, A. A.; Yates, J. R.; Souza, I.; Vanderbilt, D. Maximally localized Wannier functions: Theory and applications. *Rev. Mod. Phys.* **2012**, *84*, 1419–1475.

(52) Zicovich-Wilson, C. M.; Dovesi, R.; Saunders, V. R. A general method to obtain well localized Wannier functions for composite energy bands in linear combination of atomic orbital periodic calculations. *J. Chem. Phys.* **2001**, *115*, 9708–9719.

(53) Ziółkowski, M.; Jansík, B.; Kjærgaard, T.; Jørgensen, P. Linear scaling coupled cluster method with correlation energy based error control. *J. Chem. Phys.* **2010**, *133*, 014107.

(54) Kristensen, K.; Ziółkowski, M.; Jansík, B.; Kjærgaard, T.; Jørgensen, P. A Locality Analysis of the Divide–Expand–Consolidate Coupled Cluster Amplitude Equations. *J. Chem. Theory Comput.* **2011**, *7*, 1677–1694.

(55) Høyvik, I.-M.; Kristensen, K.; Jansik, B.; Jørgensen, P. The divide-expand-consolidate family of coupled cluster methods. *J. Chem. Phys.* **2012**, *136*, 014105.

(56) Eriksen, J. J.; Baudin, P.; Ettenhuber, P.; Kristensen, K.; Kjærgaard, T.; Jørgensen, P. Linear-Scaling Coupled Cluster with Perturbative Triple Excitations: The Divide–Expand–Consolidate CCSD(T) Model. *J. Chem. Theory Comput.* **2015**, *11*, 2984–2993.

(57) Baudin, P.; Ettenhuber, P.; Reine, S.; Kristensen, K.; Kjærgaard, T. Efficient linear-scaling second-order Møller-Plesset perturbation theory: The divide–expand–consolidate RI-MP2 model. *J. Chem. Phys.* **2016**, *144*, 054102.

(58) Kjærgaard, T.; Baudin, P.; Bykov, D.; Eriksen, J. J.; Ettenhuber, P.; Kristensen, K.; Larkin, J.; Liakh, D.; Pawłowski, F.; Vose, A.; Wang, Y. M.; Jørgensen, P. Massively

parallel and linear-scaling algorithm for second-order Møller–Plesset perturbation theory applied to the study of supramolecular wires. *Comput. Phys. Commun.* **2017**, *212*, 152 – 160.

(59) Kristensen, K.; Jørgensen, P.; Jansík, B.; Kjærgaard, T.; Reine, S. Molecular gradient for second-order Møller-Plesset perturbation theory using the divide-expand-consolidate (DEC) scheme. *J. Chem. Phys.* **2012**, *137*, 114102.

(60) Kristensen, K.; Kjærgaard, T.; Høyvik, I.-M.; Ettenhuber, P.; Jørgensen, P.; Jansik, B.; Reine, S.; Jakowski, J. The divide–expand–consolidate MP2 scheme goes massively parallel. *Mol. Phys.* **2013**, *111*, 1196–1210.

(61) Ettenhuber, P.; Baudin, P.; Kjærgaard, T.; Jørgensen, P.; Kristensen, K. Orbital spaces in the divide-expand-consolidate coupled cluster method. *J. Chem. Phys.* **2016**, *144*, 164116.

(62) Høyvik, I.-M.; Jørgensen, P. Characterization and Generation of Local Occupied and Virtual Hartree–Fock Orbitals. *Chem. Rev.* **2016**, *116*, 3306–3327.

(63) Bert, A.; Llunell, M.; Dovesi, R.; Zicovich-Wilson, C. M. Electronic structure characterization of six semiconductors through their localized Wannier functions. *Phys. Chem. Chem. Phys.* **2003**, *5*, 5319–5325.

(64) Høyvik, I.-M.; Olsen, J.; Jørgensen, P. Generalising localisation schemes of orthogonal orbitals to the localisation of non-orthogonal orbitals. *Mol. Phys.* **2017**, *115*, 16–25.

(65) Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalleri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D'Arco, P.; Noel, Y.; Causa, M.; Rerat, M.; Kirtman, B. CRYSTAL14: A Program for the *Ab Initio* Investigation of Crystalline Solids. *Int. J. Quantum Chem.* **2014**, *114*, 1287–1317.

(66) Foster, J.; Boys, S. F. Canonical Configuration Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300–302.

(67) Jansík, B.; Høst, S.; Kristensen, K.; Jørgensen, P. Local orbitals by minimizing powers of the orbital variance. *J. Chem. Phys.* **2011**, *134*, 194104.

(68) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.

(69) Boughton, J. W.; Pulay, P. Comparison of the Boys and Pipek-Mezey localizations in the local correlation approach and automatic virtual basis selection. *J. Comput. Chem.* **1993**, *14*, 736–740.

(70) Hampel, C.; Werner, H.-J. Local treatment of electron correlation in coupled cluster theory. *J. Chem. Phys.* **1996**, *104*, 6286–6297.

(71) Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, 034106.

(72) Häser, M.; Ahlrichs, R. Improvements on the direct SCF method. *J. Comput. Chem.* **1989**, *10*, 104–111.

(73) Valeev, E. F. Libint: A library for the evaluation of molecular integrals of many-body operators over Gaussian functions, version 2.4.2. `http://libint.valeyev.net/`, 2017.

(74) Chapman, B.; Jost, G.; van der Pas, R. *Using OpenMP. Portable Shared Memory Parallel Programming*; The MIT Press, 2008.

(75) iMatrix Corporation, ØMQ Distributed Messaging. `http://zeromq.org`, 2014; Accessed 2017-11-25.

(76) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(77) Feller, D. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.* **1996**, *17*, 1571–1586.

(78) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. Basis Set Exchange: A Community Database for Computational Sciences. *J. Chem. Inf. Mod.* **2007**, *47*, 1045–1052.

(79) Environmental Molecular Sciences Laboratory, Basis Set Exchange Portal. `https://bse.pnl.gov/bse/portal`, 2017; Accessed 2017-11-25.

(80) Bunn, C. W. The crystal structure of ethylene. *Trans. Faraday Soc.* **1944**, *40*, 23–25.

(81) Kalugina, Y. N.; Cherepanov, V. N.; Buldakov, M. A.; Zvereva-Loëte, N.; Boudon, V. Theoretical investigation of the ethylene dimer: Interaction energy and dipole moment. *J. Comput. Chem.* **2012**, *33*, 319–330.

(82) Program cryapi_inp. `http://www.crystal.unito.it/documentation.php`, 2017; Accessed 2017-11-25.

(83) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Ekström, U.; Enevoldsen, T.; Eriksen, J. J.; Ettenhuber, P.; Fernández, B.; Ferrighi, L.; Fliegl, H.; Frediani, L.; Hald, K.; Halkier, A.; Hättig, C.; Heiberg, H.; Helgaker, T.; Hennum, A. C.; Hettema, H.; Hjertenæs, E.; Høst, S.; Høyvik, I.-M.; Iozzi, M. F.; Jansík, B.; Jensen, H. J. Aa.; Jonsson, D.; Jørgensen, P.; Kauczor, J.; Kirpekar, S.; Kjærgaard, T.; Klopper, W.; Knecht, S.; Kobayashi, R.; Koch, H.; Kongsted, J.; Krapp, A.; Kristensen, K.; Ligabue, A.; Lutnæs, O. B.; Melo, J. I.; Mikkelsen, K. V.; Myhre, R. H.; Neiss, C.;

Nielsen, C. B.; Norman, P.; Olsen, J.; Olsen, J. M. H.; Osted, A.; Packer, M. J.; Pawlowski, F.; Pedersen, T. B.; Provasi, P. F.; Reine, S.; Rinkevicius, Z.; Ruden, T. A.; Ruud, K.; Rybkin, V. V.; Sałek, P.; Samson, C. C. M.; de Merás, A. S.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sneskov, K.; Steindal, A. H.; Sylvester-Hvid, K. O.; Taylor, P. R.; Teale, A. M.; Tellgren, E. I.; Tew, D. P.; Thorvaldsen, A. J.; Thøgersen, L.; Vahtras, O.; Watson, M. A.; Wilson, D. J. D.; Ziolkowski, M.; Ågren, H. The Dalton quantum chemistry program system. *WIREs Comput. Mol. Sci.* **2014**, *4*, 269–284.

(84) `http://daltonprogram.org`, 2015; LSDalton, a linear scaling molecular electronic structure program, Release Dalton2016.0.

(85) Maschio, L. 2017; Private communication.

(86) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley and Sons, Ltd: Chichester, 2000.

(87) The Scipy community, scipy.interpolate.splrep. `https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.interpolate.splrep.html`, Accessed 2018-01-05.

(88) Dierckx, P. An algorithm for smoothing, differentiation, and integration of experimental data using spline functions. *J. Comput. Appl. Math.* **1975**, *1*, 165–184.

(89) Jónsson, E. Ö.; Lehtola, S.; Puska, M.; Jónsson, H. Theory and Applications of Generalized Pipek–Mezey Wannier Functions. *J. Chem. Theory Comput.* **2017**, *13*, 460–474.

(90) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem. Phys. Lett.* **1993**, *208*, 359–363.