JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

A Locality Analysis of the Divide−Expand−Consolidate Coupled Cluster Amplitude Equations

Kasper Kristensen,[†,*] Marcin Ziółkowski,[‡,†] Branislav Jansík,[†] Thomas Kjærgaard,[§,†] and Poul Jørgensen[†]

[†]Lundbeck Foundation Center for Theoretical Chemistry, Department of Chemistry, University of Aarhus, DK-8000 Århus C, Denmark

ABSTRACT: We present a thorough locality analysis of the divide−expand−consolidate amplitude equations for second-order Møller−Plesset perturbation theory and the coupled cluster singles doubles (CCSD) model, which demonstrates that the amplitude equations are local when expressed in terms of a set of local occupied and local unoccupied Hartree−Fock orbitals, such as the least-change molecular basis. The locality analysis thus shows that a CC calculation on a large molecular system may be carried out in terms of CC calculations on small orbital fragments of the total molecular system, where the sizes of the orbital fragment spaces are determined in a black box manner to ensure that the CC correlation energy is calculated to a preset energy threshold. A practical implementation of the locality analysis is described, and numerical results are presented, which demonstrate that both the orbital fragment sizes and the relative energy error compared to a full CC calculation are independent of the molecular system size.

1. INTRODUCTION

A coupled cluster (CC) calculation[1] starts out with a mean field Hartree−Fock (HF) calculation,[2,3] which gives the total HF energy and a set of occupied and unoccupied HF molecular orbitals. The CC wave function is expanded in the HF orbital basis, and the CC amplitudes are determined by solving sets of nonlinear amplitude equations. From the CC amplitudes the correlation energy is determined. The total CC energy is the sum of the HF energy and the correlation energy.

The mean field HF calculation gives a good description of the long-range interaction between the electrons, while the interaction for shorter distances is described with less accuracy. The major task of a CC calculation is therefore to describe these short-ranged electron−electron interactions. Using a simplistic physical picture, this interaction may be expressed in terms of the very short-ranged interaction leading to coulomb holes in the wave function and the longer ranged interaction leading to dispersion effects. Both these effects describe local phenomena.

Standard CC calculations are expressed in the canonical HF basis which is a nonlocal basis, where the individual HF orbitals extend over the whole molecular system. The description of local phenomena using a nonlocal basis will inherently lead to a scaling wall for the standard CC calculations. To circumvent this scaling wall, it has been attempted to express the CC wave function in a basis of local HF orbitals.[4]

Local occupied HF orbitals have been obtained by localizing the canonical HF orbitals using various standard techniques.[5−7] However, none of these techniques have been able to yield a set of local unoccupied HF orbitals.[8] Projected atomic orbitals (PAOs), where the occupied HF orbital space is projected out of the atomic basis, have been used to span the unoccupied HF orbital space. The PAOs constitute a redundant set, but a more severe drawback in a local wave function context is that the locality of the PAOs is much less pronounced than the locality of the localized occupied HF orbitals.

The local correlation wave function method development was pioneered by Pulay[4] and Saebø and Pulay,[9] and the local CC method of Hampel and Werner[10] and Schütz and Werner[11−13] constitutes a prominent early contribution. Many other local CC methods have been proposed,[14−25] including atomic orbital-based CC,[14−16] the natural linear scaling approach,[18] the cluster-in-molecule approach,[19−21] the the divide-and-conquer approach,[22] the fragment molecular orbital approach,[23] and the incremental scheme.[24,25] For the simplest correlation method, second-order Møller−Plesset perturbation theory (MP2), other approaches have also been proposed where linear scaling have been obtained[26] with Laplace transformations of the energy denominator and applying state-of-the-art integral screening techniques.[27−29]

We have recently described how a local HF orbital basis—the least-change molecular (LCM) basis[30]—can be obtained for both the occupied and the unoccupied HF orbital spaces. In particular, we have found that the locality of the unoccupied HF orbitals is not very different from the locality of the occupied HF orbitals for standard basis sets, such as cc-pVDZ and cc-pVTZ. By applying this local HF basis and restructuring the CC equations, we have developed the linear-scaling and embarrassingly parallel divide−expand−consolidate (DEC) CC method.[31] The DEC method is on par with the standard CC method in the sense that the precision of the correlation energy is defined by a preset threshold. Specifically, in a standard CC calculation the precision is specified by a preset threshold for the residual norm for the CC amplitude equations, and in a DEC calculation the precision is specified by a preset threshold for the atomic fragment energies, which directly reflects the accuracy in the total correlation energy (provided that a tight residual threshold is used in the amplitude equations for the fragment calculations). In short, the accuracy in standard CC and DEC CC is defined prior to the calculation.

In the Laplace integral screening AO-MP2 approach the error compared to a full MP2 calculation can also be controlled in a precise manner,[27−29] while in existing local CC methods the standard CC correlation energy is an asymptotic limit, which is

Received: February 17, 2011
Published: April 18, 2011

ACS Publications © 2011 American Chemical Society

obtained when all thresholds are removed. However, in practice, ad hoc approximations to the standard CC method are introduced in the local CC methods, where the effects of these approximations on the correlation energy and the molecular properties are not known. The approximations include a priori assignments of local orbital spaces, which introduce local correlation errors, e.g., when the completeness criterion of Boughton and Pulay[32] is used to assign a priori fixed orbital excitation spaces to localized occupied HF orbitals. Some approximations also include nonphysical bond cuts of the molecular system. At the end of a local CC calculation, where ad hoc approximations have been introduced, the precision of the calculation compared to a full CC calculation is in general unknown.

One could argue that compared to the large errors in CC calculations due to the basis set incompleteness and the approximate CC wave function model, the approximations that are introduced in the local CC methods are acceptable. However, in chemistry we are typically interested in the energy differences for molecular conformers, for example, reaction enthalpies, interaction energies of van der Waals complexes, excitation energies, and geometrical energy derivatives. These energy differences are small compared to total energies, and to obtain a reliable and accurate description of these small differences, it is necessary to know the errors introduced by the local approximations.

Comparing the DEC approach to existing local CC methods, we conclude that using the DEC approach, the accuracy in the energy is defined prior to the calculation, and the sizes of the orbital spaces are adjusted during the calculation according to this accuracy. In this way the DEC scheme ensures that the standard CC energy is reproduced to the requested accuracy using as small orbital spaces as possible. Existing local CC approaches in general lack the flexibility to allow the sizes of the orbital spaces to adjust to an energy threshold, and therefore the error compared to the standard CC energy is in general unknown.

The DEC energy equations have previously been presented.[31] In this paper we perform a thorough locality analysis of the MP2 and CC singles doubles (CCSD) amplitude equations, which demonstrate that the amplitude equations are indeed local when expressed in the LCM basis,[30] where both the occupied and the virtual orbitals are local. The locality analysis shows that a CC calculation on a molecular system may be carried out in terms of CC calculations on small orbital fragments of the total molecular system, when the fragmentation of the orbital spaces is carried out as in the DEC CC method. We also present a practical implementation of the DEC method. In particular, we discuss the removal of orbital tail coefficients to obtain more local molecular orbitals (and thus smaller atomic orbital spaces for evaluating two-electron integrals), how orbital spaces are determined according to a predefined energy accuracy, and how to avoid wave function superposition errors when pair interaction energies are calculated. Numerical results are given, which demonstrate that the sizes of the fragment orbital spaces are independent of the molecular system size—in agreement with the theoretical locality analysis—and that the standard CC correlation energy can be recovered to a predefined accuracy using relatively small orbital fragment spaces.

In the next section we derive the basic equations of the DEC model. In Sections 3 and 4 we perform a locality analysis for the MP2 and CCSD amplitude equations. In Section 5 the computational scaling in DEC calculations is discussed, and Section 6 contains implementation details for the DEC method. In Section 7 we present some illustrative DEC calculations. Finally, in Section 8 we give a short summary of the DEC model and some concluding remarks.

2. DEC MODEL

2.1. CC Energy. For a closed shell molecule the total energy E_{CC} for a CC wave function model may be expressed as[33]

$$E_{CC} = E_{HF} + E_{corr} \quad (1)$$

where E_{HF} is the HF total energy, and the correlation energy E_{corr} is obtained as

$$E_{corr} = \sum_{ijab} (t_{ij}^{ab} + t_i^a t_j^b)(2g_{iajb} - g_{ibja}) \quad (2)$$

where t_i^a and t_{ij}^{ab} are the cluster singles and doubles amplitudes, respectively. Indices $i, j, \ldots$ refer to occupied HF orbitals and $a, b, \ldots$ to unoccupied HF orbitals, and g_{iajb} is a two-electron integral in the HF orbital basis using the Mulliken notation. The correlation energy expression in eq 2 applies to all standard CC models. The singles contribution vanishes for MP2.

In a conventional CC calculation the CC wave function is expanded in the nonlocal canonical HF orbital basis, and therefore all integrals and cluster amplitudes in E_{corr} are nonvanishing. The evaluation of E_{corr} therefore has a fourth power scaling in system size. When a local HF orbital basis is used the correlation energy can be expressed in terms of orbital spaces which reference only local parts of the molecular system. We describe in the following sections how this may be accomplished.

2.2. CC Energy for Local HF Orbitals. For each local HF orbital the Mulliken charge is determined, and the HF orbital is assigned to the atomic site with the largest Mulliken charge. In this way each atomic site gets assigned a set of local occupied and local unoccupied HF orbitals. The set of occupied HF orbitals assigned to atomic site P is denoted $\underline{P}$, and the set of unoccupied HF orbitals assigned to atomic site P is denoted $\overline{\mathrm{P}}$. A cartoon illustrating how a model one-dimensional molecular system is divided into atomic sites and assigned a set of occupied and unoccupied HF orbitals is given in Figure 1A. As a specific example we present the orbital assignment for the $C_{14}H_2$ molecule (LCM molecular orbitals using a cc-pVDZ basis set) in Figure 1B. (In Section 7.1.1 we elaborate on the assignment of orbitals in Figure 1B.) In Table 1 we give an overview of the notation for orbital spaces and energies used for describing the DEC model.

Having assigned the local HF orbitals to atomic sites, the correlation energy in eq 2 can be expressed in terms of atomic fragment energies E_P and atomic pair fragment energies E_{PQ}:

$$E_P = \sum_{\substack{ij \in \underline{P} \\ ab}} (t_{ij}^{ab} + t_i^a t_j^b)(2g_{iajb} - g_{ibja}) \quad (3)$$

$$E_{PQ} = \sum_{\substack{ij \in \underline{P} \cup \underline{Q} \\ ab}} (t_{ij}^{ab} + t_i^a t_j^b)(2g_{iajb} - g_{ibja}) \quad (4)$$

giving

$$E_{corr} = \sum_P E_P + \sum_{P>Q} \Delta E_{PQ} \quad (5)$$

where the sum runs over all atomic sites, and the pair interaction energy ΔE_{PQ} is given as

$$\Delta E_{PQ} = E_{PQ} - E_P - E_Q \quad (6)$$

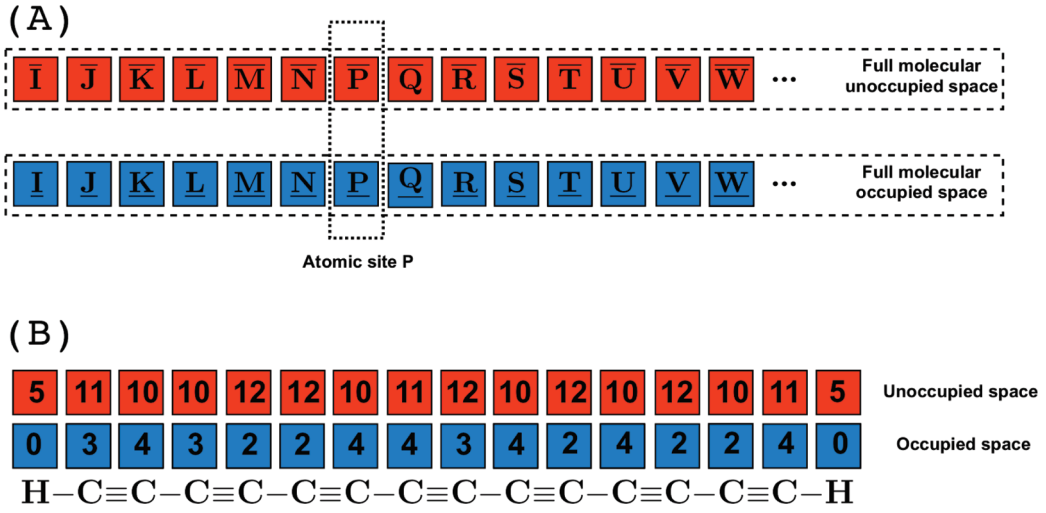

Figure 1. (A) Molecular system divided into atomic sites I, J, ..., P, ..., where each site has been assigned a set of occupied (blue) and unoccupied (red) HF orbitals. (B) Example of orbital assignments for the $C_{14}H_2$ molecule (LCM molecular orbitals using a cc-pVDZ basis set). The values above each atom denote the number of occupied and virtual orbitals assigned to that atom.

Table 1. Overview of Notation for Atomic Sites, Energies, And Orbitals Spaces

symbol	short description
P, Q, R, ...	labels for atomic sites
E_P	atomic fragment energy
E_{PQ}	atomic pair fragment energy
ΔE_{PQ}	pair interaction energy
$\mathscr{E}_P$	energy orbital space (EOS) for evaluating E_P
$\mathscr{A}_P$	amplitude orbital space (AOS) for calculating CC amplitudes
$\{\mathscr{A}_P\}$	atomic fragment extent where MO coefficients are expanded
$\underline{P}$	set of occupied orbitals assigned to site P (occupied orbitals in $\mathscr{E}_P$)
$\overline{P}$	set of unoccupied orbitals assigned to site P
$[\underline{P}]$	set of occupied orbitals local to P
$[\overline{P}]$	set of unoccupied orbitals local to P (unoccupied orbitals in $\mathscr{E}_P$)
$[\underline{P}]_2$	set of occupied orbitals local to $[\underline{P}]$ (including $[\underline{P}]$)
$[\overline{P}]_2$	set of unoccupied orbitals local to $[\overline{P}]$ (including $[\overline{P}]$)
$\mathscr{B}$	occupied buffer orbital space
$\mathscr{B}_{\overline{P}}$	unoccupied buffer orbital space

No approximations have been made in eq 5, where the sum over the two occupied orbital indices in eq 2 has been replaced by sums over the occupied orbitals of the atomic fragments and atomic pair fragments. The first term in eq 5 describes the coulomb holes in the wave function, while the second term for larger PQ distances mainly describes dispersion effects.

In the local HF orbital basis the integral g_{iajb} is nonvanishing only if the orbital pair indices ia refer to the same atomic site or to neighboring atomic sites and similarly for jb. Therefore the integral g_{iajb} is nonvanishing only if $i \in \underline{P}$ and $a \in [\overline{P}]$, where $[\overline{P}]$ refers to the unoccupied HF orbital space local to $\overline{P}$ (including $\overline{P}$) and similarly for jb. The requirements for a nonvanishing integral may thus be summarized as

$$g_{iajb} : i \in \underline{P}, \quad a \in [\overline{P}], \quad j \in \underline{Q}, \quad b \in [\overline{Q}] \tag{7}$$

When the atomic fragment energy E_P in eq 3 is calculated, both occupied indices belong to atomic site P [$P = Q$ in eq 7]. Since the integral $(2g_{iajb} - g_{ibja})$ in E_P is nonvanishing only if $a,b \in [\overline{P}]$, we may restrict the virtual index summation in eq 3 and evaluate E_P as

$$E_P = \sum_{\substack{ij \in \underline{P} \\ ab \in [\overline{P}]}} (t_{ij}^{ab} + t_i^a t_j^b)(2g_{iajb} - g_{ibja}) \tag{8}$$

where the summation over the unoccupied indices has been restricted to atomic sites, which are local to P. Similarly, the locality of the integral $(2g_{iajb} - g_{ibja})$ in eq 4 implies that the atomic pair fragment energy E_{PQ} may be evaluated as

$$E_{PQ} = \sum_{\substack{ij \in \underline{P} \cup \underline{Q} \\ ab \in [\overline{P}] \cup [\overline{Q}]}} (t_{ij}^{ab} + t_i^a t_j^b)(2g_{iajb} - g_{ibja}) \tag{9}$$

where the unoccupied index summation has been restricted to the union of unoccupied spaces which are local to $\overline{P}$ and $\overline{Q}$ using eq 7. The total orbital space required for evaluating E_{PQ} is thus the union of orbital spaces for evaluating E_P and E_Q.

Disregarding singles the atomic fragment energy E_P depends on the doubles amplitudes:

$$t_{ij}^{ab} : i,j \in \underline{P}, \quad a,b \in [\overline{P}] \tag{10}$$

The orbital space defined by eq 10 will be denoted the atomic fragment energy orbital space (EOS) $\mathscr{E}_P$. The atomic pair fragment energy E_{PQ} depends on the doubles amplitudes

$$t_{ij}^{ab} : i,j \in \underline{P} \cup \underline{Q}, \quad a,b \in [\overline{P}] \cup [\overline{Q}] \tag{11}$$

and the orbital space defined by eq 11 is denoted the atomic pair fragment EOS $\mathscr{E}_{PQ} = \mathscr{E}_P \cup \mathscr{E}_Q$.

A CC calculation carried out in terms of atomic fragment and pair fragment calculations will be denoted a DEC calculation because it follows the following procedure:

- Divide the orbital space among the atomic sites.
- Expand the orbital space of the individual atomic fragments to obtain converged fragment energies.

- Consolidate the fragment energies according to eq 5 to get the correlation energy for the full molecular system.

3. DEC-MP2 AMPLITUDE EQUATIONS

In this section the MP2 amplitude equations are analyzed in the local HF basis to understand how the atomic fragment and the atomic pair fragment energies can be determined from calculations referencing only small fragments of the full molecular orbital space. In the next section a similar analysis will be performed for the CCSD amplitude equations.

The locality analysis of the amplitude equations demonstrates that a full MP2 calculation can be carried out in terms of small independent fragment calculations. The actual fragment sizes are determined during the DEC calculation in order to ensure that the fragment energies are determined to a given threshold, as will be discussed in Section 6.2.

3.1. MP2 Amplitude Equations. Using standard notation[33] the MP2 amplitude equations may be expressed as

$$\Omega_{aibj} = \left\langle \begin{matrix} \overline{ab} \\ ij \end{matrix} \middle| \hat{H} \middle| \mathrm{HF} \right\rangle + \left\langle \begin{matrix} \overline{ab} \\ ij \end{matrix} \middle| \left[\hat{F}, \hat{T}_2\right] \middle| \mathrm{HF} \right\rangle = 0 \quad (12)$$

where $\hat{H}$ is the Hamiltonian:

$$\hat{H} = \sum_{rs} h_{rs} E_{rs} + \frac{1}{2} \sum_{rstu} g_{rstu} (E_{rs} E_{tu} - \delta_{st} E_{ru}) \quad (13)$$

$\hat{T}_2$ is the cluster doubles operator:

$$\hat{T}_2 = \frac{1}{2} \sum_{aibj} t_{ij}^{ab} E_{ai} E_{bj} \quad (14)$$

and $\left\langle \begin{matrix} \overline{ab} \\ ij \end{matrix} \right|$ is a doubles bra state in the biorthonormal basis. $\hat{F}$ is the Fock operator:

$$\hat{F} = \sum_{pq} F_{pq} E_{rs}$$

where the Fock matrix elements are given as

$$F_{pq} = h_{pq} + \sum_{i} (2 g_{pqii} - g_{piiq}) \quad (15)$$

The Fock matrix has vanishing matrix elements between occupied and unoccupied HF orbitals and therefore has a block diagonal structure:

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}^{oo} & 0 \\ 0 & \mathbf{F}^{vv} \end{pmatrix} \quad (16)$$

where $\mathbf{F}^{oo}$ and $\mathbf{F}^{vv}$ denote the occupied−occupied and unoccupied−unoccupied blocks of the Fock matrix, respectively. The MP2 amplitude equation in eq 12 may be evaluated as

$$\Omega_{aibj}^{\mathrm{MP2}} = g_{aibj} + \sum_{c} t_{ij}^{cb} F_{ac} + \sum_{c} t_{ij}^{ac} F_{bc} - \sum_{k} t_{kj}^{ab} F_{ki} - \sum_{k} t_{ik}^{ab} F_{kj} = 0 \quad (17)$$

and becomes a set of linear equations which determine the cluster amplitudes.

In the canonical HF basis where $\mathbf{F}^{oo}$ and $\mathbf{F}^{vv}$ are diagonal matrices with orbital energies on the diagonal ($F_{ij}^{oo} = \delta_{ij}\varepsilon_j$, $F_{ab}^{vv} = \delta_{ab}\varepsilon_a$), the amplitude equations in eq 17 decouple, and the solution to eq 17 becomes trivial:

$$t_{ij}^{ab} = -g_{aibj} (\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j)^{-1} \quad (18)$$

However, in the canonical basis, where the HF orbitals are nonlocal, all integrals g_{aibj} and therefore also all amplitudes t_{ij}^{ab} are, in general, nonvanishing. The canonical basis is therefore not suited for a DEC calculation.

3.2. Orbital Spaces for Evaluating E_P and E_{PQ} for MP2. For a local HF basis the correlation energy is obtained as a sum of the atomic fragment and the pair fragment energies in eqs 8 and 9, which are determined from the amplitudes of the EOSs defined by eqs 10 and 11. We describe below how the amplitudes of the EOS can be determined from MP2 calculations, which reference only a small fragment of the full molecular orbital space. The selection of the orbital spaces for the small orbital fragment calculations is performed based on a locality analysis of the MP2 amplitude equations in a local HF basis. Since we are interested in determining the EOS amplitudes in eqs 10 and 11, we consider the propagation of the amplitude equations in eq 17 for $i \in \underline{P}$ and $j \in \underline{Q}$.

The amplitude equations in eq 17 constitute a set of linear equations, which can be ordered to give a positive definite coefficient matrix:

$$\sum_{c} t_{ij}^{cb} F_{ac} + \sum_{c} t_{ij}^{ac} F_{bc} - \sum_{k} t_{kj}^{ab} F_{ki} - \sum_{k} t_{ik}^{ab} F_{kj} = -g_{aibj} \quad (19)$$

Equation 19 may be solved using standard iterative algorithms, such as the conjugate gradient or conjugate residual methods. When solving eq 19 using one of these methods, a new direction for the $(n + 1)$'th iteration is determined from the residual containing the amplitudes of the n'th iteration $t_{ij}^{n,ab}$:

$$t_{ij}^{n+1,ab} = t_{ij}^{n,ab} + R_{aibj}^{n,\mathrm{MP2}}(t^n) \quad (20)$$

where $R_{aibj}^{n,\mathrm{MP2}}(t^n)$ is the residual of eq 19 for iteration n:

$$\begin{aligned} R_{aibj}^{n,\mathrm{MP2}}(t^n) &= -\Omega_{aibj}^{n,\mathrm{MP2}}(t^n) \\ &= -g_{aibj} - \sum_{c} t_{ij}^{n,cb} F_{ac} - \sum_{c} t_{ij}^{n,ac} F_{bc} + \sum_{k} t_{kj}^{n,ab} F_{ki} \\ &\quad + \sum_{k} t_{ik}^{n,ab} F_{kj} \end{aligned} \quad (21)$$

In the conjugate gradient or the conjugate residual methods, a line search is performed along the residual direction. Furthermore, the convergence may be improved by preconditioning the linear equations. Both line search and preconditioning will affect the convergence rate of the algorithm, but they will have no effect on how new orbital spaces are introduced in each iteration, starting from $i \in \underline{P}$ and $j \in \underline{Q}$. Since the focus in this analysis is on the propagation of orbital spaces, we choose to carry out the analysis using the simplified algorithm outlined above without line search and preconditioning. We note that in practice we use the conjugate residual algorithm and precondition the equation using the diagonal elements of the Fock matrix.

Before carrying out the locality analysis it is instructive to demonstrate that for a diagonal Fock matrix the simple algorithm outlined above converges to the trivial solution in eq 18. Schematically, eq 19 may then be written as

$$\varepsilon t = -g \quad (22)$$

where $\varepsilon = \varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j$ is positive. The residual in eq 21 now becomes

$$R^n = -g - \varepsilon t^n \quad (23)$$

In each iteration the t amplitudes are updated as described in eq 20:

$$t^{n+1} = t^n + R^n = -g + (1-\varepsilon)t^n \tag{24}$$

where we have inserted eq 23. Starting out with $t^1 = 0$ we thus obtain

$$t^2 = -g \tag{25}$$

$$t^3 = -g[1 + (1-\varepsilon)] \tag{26}$$

$$t^4 = -g[1 + (1-\varepsilon) + (1-\varepsilon)^2] \tag{27}$$

... (28)

$$t^{n+1} = -g[1 + (1-\varepsilon) + (1-\varepsilon)^2 + \ \ldots \ + (1-\varepsilon)^{n-1}] \tag{29}$$

Noting that for $0 > \varepsilon > 1$, we may expand $1/\varepsilon$ as

$$\frac{1}{\varepsilon} = 1 + (1-\varepsilon) + (1-\varepsilon)^2 + \ldots \ (0 < \varepsilon < 1) \tag{30}$$

it is seen that the amplitude update in eq 20 indeed converges to solution of eq 22:

$$t^{\infty} = -g[1 + (1-\varepsilon) + (1-\varepsilon)^2 + \ldots \ ] = -\frac{g}{\varepsilon} \tag{31}$$

Let us now return to analyzing the propagation of the amplitude equations in eq 19 in a local HF basis. We start out by considering $i \in \underline{P}$ and $j \in \underline{Q}$ in the first iteration to investigate how the EOS amplitudes in eqs 10 and 11 couple to other amplitudes at different stages of the iteration sequence. In particular, we will investigate the local orbital spaces entering iterations two, three, and four, i.e., we consider the local amplitude analogues of the diagonal expressions in eqs 25−27. First we note that the coupling of amplitudes in eq 19 occurs via the off-diagonal elements in $\mathbf{F}^{oo}$ and $\mathbf{F}^{vv}$, in contrast to the simple diagonal case above, where the amplitudes are uncoupled. The Fock matrix in the local basis in eq 14 is, however, diagonally dominant in the sense that the largest elements of $\mathbf{F}^{oo}$ and $\mathbf{F}^{vv}$ occur for pair indices ij or ab, where the orbital indices refer to the same atomic site. Furthermore, the magnitudes of the off-diagonal Fock matrix elements decrease with an increasing distance between the atomic sites. Non-negligible off-diagonal Fock matrix elements are therefore encountered only if

$$F^{oo}_{kl}; k \in \underline{P}, \ l \in [\underline{P}] \quad \text{or} \quad k \in [\underline{P}], \ l \in \underline{P} \tag{32}$$

$$F^{vv}_{ab}; a \in \overline{P}, \ b \in [\overline{P}] \quad \text{or} \quad a \in [\overline{P}], \ b \in \overline{P} \tag{33}$$

The iterative algorithm is started by setting the amplitudes $t^1 = 0$, which, using eq 21, gives the residual

$$R^{1,\mathrm{MP2}}_{aibj}(t^1) = -g_{aibj} \tag{34}$$

For $i \in \underline{P}$ and $j \in \underline{Q}$ the integrals g_{aibj} are nonvanishing only if $a \in [\overline{P}]$ and $b \in [\overline{Q}]$, see eq 7. In the first iteration we therefore consider only the residuals for the orbital space:

$$R^{1,\mathrm{MP2}}_{aibj}(t^1) : i \in \underline{P}, \ a \in [\overline{P}], \ j \in \underline{Q}, \ b \in [\overline{Q}] \tag{35}$$

which is identical to the EOS orbital space $\mathcal{S}_{\mathrm{PQ}}$ in eq 11. The amplitudes of the second iteration therefore become

$$t^{2,ab}_{ij} = t^{1,ab}_{ij} + R^{1,\mathrm{MP2}}_{aibj}(t^1) = R^{1,\mathrm{MP2}}_{aibj}(t^1) \tag{36}$$

and are nonvanishing only for the EOS:

$$t^{2,ab}_{ij} : i \in \underline{P}, \ a \in [\overline{P}], \ j \in \underline{Q}, \ b \in [\overline{Q}] \tag{37}$$

When the residual in the EOS vanishes, the amplitudes that determine E_{PQ} are converged. Carrying out an iterative procedure the convergence of the residual in the EOS is fast due to the diagonal dominance of the Fock matrix. New orbital spaces are introduced during the iterative procedure, however the most important effect for the convergence of the residual for the EOS is the relaxation of the amplitudes for the orbital spaces already considered.

In the second iteration the orbital space which interacts directly with the EOS is determined. The residual in the second iteration becomes

$$R^{2,\mathrm{MP2}}_{aibj}(t^2) = -g_{aibj} - \sum_{c \in [\overline{P}]} t^{2,cb}_{ij} F_{ac} - \sum_{c \in [\overline{Q}]} t^{2,ac}_{ij} F_{bc} + \sum_{k \in \underline{P}} t^{2,ab}_{kj} F_{ki} + \sum_{k \in \underline{Q}} t^{2,ab}_{ik} F_{kj} \tag{38}$$

For the second term in eq 38, the summation index is restricted to $c \in [\overline{P}]$ due to the restrictions on the amplitudes in eq 37. The second term thus gives a nonvanishing residual for $i \in \underline{P}$, $a \in [\overline{P}]_2$, $j \in \underline{Q}$, $b \in [\overline{Q}]$, where we have used eq 33 to restrict the a index such that $[\overline{P}]_2$ refers to unoccupied orbitals assigned to atoms that are local to $[\overline{P}]$ (including $[\overline{P}]$ itself). (Similarly, $[\overline{P}]_3$ is the unoccupied space local to $[\overline{P}]_2$ and so on.) The integral contribution $(-g_{aibj})$ to the residual is vanishing for the extended orbital space $i \in \underline{P}$, $a \in [\overline{P}]_2 - [\overline{P}]$, $j \in \underline{Q}$, $b \in [\overline{Q}]$, see eq 7, and therefore the residual in this extended orbital space is small. The third term may be treated similarly as the second term. For the fourth term in eq 38 the summation is restricted to $k \in \underline{P}$ due to the restrictions on the amplitudes in eq 37. The fourth term therefore gives a nonvanishing residual contribution for the orbital space $i \in \underline{P}$, $a \in [\overline{P}]$, $j \in \underline{Q}$, $b \in [\overline{Q}]$, where we have also used eq 32 to restrict the i index. In this case the extended orbital space $i \in [\underline{P}] - \underline{P}$, $a \in [\overline{P}]$, $j \in \underline{Q}$, $b \in [\overline{Q}]$ has a nonvanishing integral contribution to the residual, and the residual may therefore be of the same magnitude as the residual in the first iteration. The last term in eq 38 may be treated similarly as the fourth term. In conclusion, nonvanishing residuals are obtained in the second iteration for the orbital spaces:

$$R^{2,\mathrm{MP2}}_{aibj} : i \in [\underline{P}], \ a \in [\overline{P}]_2, \ j \in [\underline{Q}], \ b \in [\overline{Q}]_2 \tag{39}$$

where residuals for the extended spaces $i \in [\underline{P}] - \underline{P}$, $a \in [\overline{P}]$, $j \in [\underline{Q}] - \underline{Q}$, $b \in [\overline{Q}]$ have both an integral and a coupling term contribution, while the residual for the extended spaces $i \in \underline{P}$, $a \in [\overline{P}]_2 - [\overline{P}]$, $j \in \underline{Q}$, $b \in [\overline{Q}]_2 - [\overline{Q}]$ only has a coupling term contribution.

The amplitudes of the third iteration becomes

$$t^{3,ab}_{ij} = t^{2,ab}_{ij} + R^{2,\mathrm{MP2}}_{aibj}(t^2) \tag{40}$$

where nonvanishing amplitudes $t^{3,ab}_{ij}$ are obtained for the nonvanishing residuals in eq 39:

$$t^{3,ab}_{ij} : i \in [\underline{P}], \ a \in [\overline{P}]_2, \ j \in [\underline{Q}], \ b \in [\overline{Q}]_2 \tag{41}$$

Iteration 2 thus introduces a first-order interaction (FOI) space to the EOS containing the amplitudes:

$$t^{3,ab}_{ij} : i \in [\underline{P}] - \underline{P}, \ a \in [\overline{P}]_2 - [\overline{P}], \ j \in [\underline{Q}] - \underline{Q}, \ b \in [\overline{Q}]_2 - [\overline{Q}] \tag{42}$$

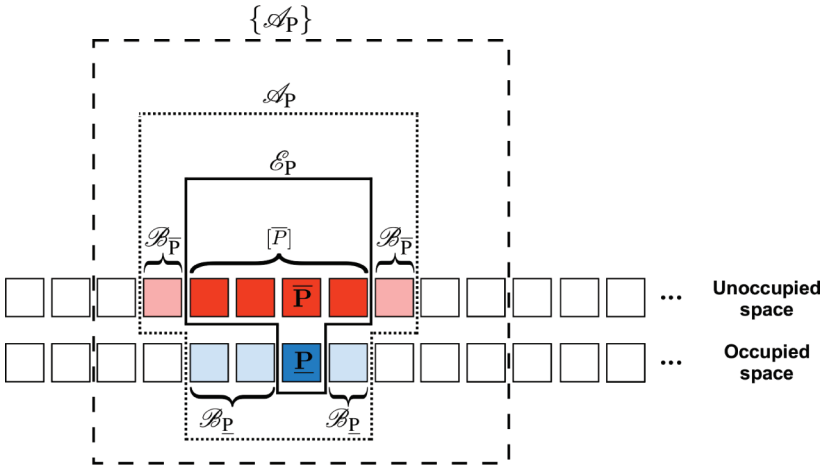

Figure 2. Atomic fragment *P*. The correlated wave function calculation is carried out using the amplitude orbital space $\mathscr{A}_P$ (dark-blue, light-blue, red, and pink markings), while the atomic fragment energy E_P is evaluated using the energy orbital space $\mathscr{E}_P$ (dark-blue and red markings). The orbitals in the molecular fragment *P* are confined to the atoms in the atomic fragment extent $\{\mathscr{A}_P\}$. Thus, two-electron integrals in the AO basis need to be calculated for atoms in the $\{\mathscr{A}_P\}$ space.

The FOI space in eq 42 is the difference between the space for the third iteration amplitudes in eq 41 and the second iteration amplitudes in eq 37. The t^3 amplitudes referencing only the indices $a \in [\overline{P}]_2 - [\overline{P}]$ and $b \in [\overline{Q}]_2 - [\overline{Q}]$ of the FOI space are small because the residuals in eq 38 have no integral contribution, while the amplitudes referencing the indices $i \in [\underline{P}] - \underline{P}$ and $j \in [\underline{Q}] - \underline{Q}$ of the FOI space may be of the same size as the amplitudes of the first iteration because the residual had an integral contribution.

Using a similar analysis for the fourth iteration, nonvanishing amplitudes may be obtained for the orbital space:

$$t_{ij}^{4,ab} : i \in [\underline{P}]_2,\ a \in [\overline{P}]_3,\ j \in [\underline{Q}]_2,\ b \in [\overline{Q}]_3 \qquad (43)$$

Iteration three thus introduces a second-order interaction (SOI) space relative to the EOS containing the amplitudes:

$$t_{ij}^{4,ab} : i \in [\underline{P}]_2 - [\underline{P}],\ a \in [\overline{P}]_3 - [\overline{P}]_2,\ j \in [\underline{Q}]_2 - [\underline{Q}],\ b \in [\overline{Q}]_3 - [\overline{Q}]_2 \qquad (44)$$

The SOI space is the difference between the spaces in eqs 43 (iteration four) and 41 (iteration three). The SOI space does not interact directly with the amplitudes of the EOS, and its effect on the residual of the EOS is therefore small, in particular for the interaction, which goes through the virtual FOI space, where the amplitudes are small. The effect of the SOI space on the amplitudes of the EOS can therefore in most cases be neglected. Continuing the iteration procedure leads to fast convergence of the amplitudes for the EOS, and the effect of new orbital spaces on the amplitudes of the EOS becomes insignificant.

The above development can be used to set up orbital spaces for the calculations which determine the amplitudes that are used for evaluating E_P and E_{PQ}. For E_P (corresponding to $P = Q$ in the above analysis) the amplitudes of eq 10 are required, and the orbital space therefore has to include the atomic fragment EOS $\mathscr{E}_P$: $i,j \in \underline{P}$ and $a,b \in [\overline{P}]$. Iteration two shows that the EOS interacts directly with the FOI space in eq 42. This interaction can be taken into account by introducing buffer spaces for the occupied ($\mathscr{B}_P = [\underline{P}] - \underline{P}$) and for the unoccupied ($\mathscr{B}_P = [\overline{P}]_2 - [\overline{P}]$) orbital spaces. The orbital space $\mathscr{A}_P$ which is used for solving the amplitude equations required to determine E_P therefore becomes

$$\mathscr{A}_P = \underline{P} + [\overline{P}] + \mathscr{B}_{\underline{P}} + \mathscr{B}_{\overline{P}} \qquad (45)$$

where the two first terms denote the spaces involved directly in the calculation of E_P (i.e., the atomic fragment EOS $\mathscr{E}_P$, see eq 8). The last two terms contain the amplitudes outside this space, which are necessary for determining the EOS amplitudes with high precision due to coupling of the amplitudes. The orbital space defined by eq 45 will be denoted the amplitude orbital space (AOS) of an atomic fragment calculation. In general, the indirect interactions on the residual of the EOS introduced in the third and higher iterations may be considered by having flexible buffer spaces, where the buffer spaces are extended until the atomic fragment energy is unaffected by further extensions. The determination of the sizes of the fragment spaces will be detailed in Section 6.2. An illustration of the orbital spaces that are in use when the atomic fragment energy E_P is evaluated is displayed in Figure 2 for a one-dimensional system. The $\{\mathscr{A}_P\}$ space will be discussed in Section 6.1.

For E_{PQ} we need the amplitudes of eq 11. The orbital space therefore has to include the atomic pair fragment EOS $\mathscr{E}_{PQ}$: $i,j \in \underline{P} \cup \underline{Q}$ and $a,b \in [\overline{P}] \cup [\overline{Q}]$. The FOI space introduces the occupied buffer

$$\mathscr{B}_{\underline{PQ}} = [\underline{P} \cup \underline{Q}] - \underline{P} \cup \underline{Q}$$

and the unoccupied buffer

$$\mathscr{B}_{\overline{PQ}} = [\overline{P}]_2 \cup [\overline{Q}]_2 - [\overline{P}] \cup [\overline{Q}]$$

orbital spaces, and the indirect interaction to the SOI and the higher order spaces is considered using the union of buffer orbital spaces for evaluating E_P and E_Q. The atomic pair fragment AOS used for evaluating E_{PQ} therefore becomes

$$\mathscr{A}_{PQ} = \mathscr{A}_P \cup \mathscr{A}_Q = \underline{P} \cup \underline{Q} + [\overline{P}] \cup [\overline{Q}] + \mathscr{B}_{\underline{PQ}} + \mathscr{B}_{\overline{PQ}} \qquad (46)$$

where the two first terms denote the spaces required for determining the atomic pair fragment energy, see eq 9, and the last two terms are the occupied and virtual buffer spaces. The orbital spaces used for the atomic pair fragment PQ calculation are thus formed as unions of orbital spaces for atomic fragments *P* and *Q*. The locality of a DEC pair fragment calculation is therefore determined by the orbital spaces that are obtained in the atomic fragment calculations.

3.3. Summary of MP2 Locality Analysis. Using an iterative algorithm analysis we have shown that the equations for the EOS amplitudes required to determine E_P and E_{PQ} in eqs 8 and 9 may be solved in small fragment orbital spaces (see eqs 45 and 46). The analysis thus demonstrates that locality may be exploited to divide a full MP2 calculation into small orbital fragment calculations, which may be carried out independently. The locality of an MP2 calculation is determined by the locality of the atomic fragment calculations, while the atomic pair fragment energies are obtained from calculations involving the union of spaces for the corresponding atomic fragment calculations. The optimal sizes of the fragment orbital spaces, i.e., $[\overline{P}]$, $\mathscr{B}_{\underline{P}}$, and $\mathscr{B}_{\overline{P}}$, are of course not known a priori. In Section 6.2 we discuss how these spaces are determined in a DEC calculation.

4. DEC-CCSD AMPLITUDE EQUATIONS

4.1. CCSD Amplitude Equations. We now consider the amplitude equations for the CCSD model for a closed-shell molecule. Following the notation of ref 33, the cluster amplitudes are obtained by solving the cluster amplitude equations:

$$\Omega_{ai} = \left\langle \begin{matrix} \overline{a} \\ i \end{matrix} \middle| \exp(-\hat{T}_2)\hat{\tilde{H}}\exp(\hat{T}_2) \middle| \mathrm{HF} \right\rangle = 0 \qquad (47)$$

and

$$\Omega_{aibj} = \left\langle \begin{matrix} \overline{ab} \\ ij \end{matrix} \middle| \exp(-\hat{T}_2)\hat{\tilde{H}}\exp(\hat{T}_2) \middle| \mathrm{HF} \right\rangle = 0 \qquad (48)$$

where $\langle \begin{matrix} \overline{a} \\ i \end{matrix} |$ and $\langle \begin{matrix} \overline{ab} \\ ij \end{matrix} |$ are single and double bra states in the biorthonormal basis. The cluster doubles operator $\hat{T}_2$ is given in eq 14, and the singles operator $\hat{T}_1$ is given as

$$\hat{T}_1 = \sum_{ai} t_i^a E_{ai} \qquad (49)$$

$\hat{\tilde{H}}$ refers to the $\hat{T}_1$ transformed Hamiltonian:

$$\hat{\tilde{H}} = \exp(-\hat{T}_1)\hat{H}\exp(\hat{T}_1)$$
$$= \hat{\tilde{H}}_1 + \hat{\tilde{H}}_2 = \sum_{rs}\tilde{h}_{rs}E_{rs} + \frac{1}{2}\sum_{rstu}\tilde{g}_{rstu}(E_{rs}E_{tu} - \delta_{st}E_{ru}) \qquad (50)$$

Above we have introduced the $\hat{T}_1$ transformed integrals:

$$\tilde{h}_{rs} = \sum_{\mu\nu} X_{\mu r}Y_{\nu s}h_{\mu\nu}, \qquad (51)$$

$$\tilde{g}_{pqrs} = \sum_{\mu\nu\sigma\omega} X_{\mu p}X_{\sigma r}Y_{\nu q}Y_{\omega s}g_{\mu\nu\sigma\omega}, \qquad (52)$$

where the transformation matrices $\mathbf{X}$ and $\mathbf{Y}$ are defined as

$$\mathbf{X} = \mathbf{C}(1 - \mathbf{t}_1^T) \qquad (53)$$

and

$$\mathbf{Y} = \mathbf{C}(1 + \mathbf{t}_1) \qquad (54)$$

where $\mathbf{C}$ is the transformation matrix from the AO basis to the local HF basis.

Table 2. Explicit Form of the CCSD Singles and Doubles Equations

contributions to singles residual

$$\Omega_{ai}^{A1} = \sum_{ckd} u_{ki}^{cd}\tilde{g}_{adkc}$$
$$\Omega_{ai}^{B1} = -\sum_{ckl} u_{kl}^{ac}\tilde{g}_{kilc}$$
$$\Omega_{ai}^{C1} = \sum_{ck} {}^{I}\tilde{F}_{kc}u_{ik}^{ac}$$
$$\Omega_{ai}^{D1} = {}^{I}\tilde{F}_{ai}$$

contribution to doubles residual

$$\Omega_{aibj}^{A2} = \tilde{g}_{aibj} + \sum_{cd} t_{ij}^{cd}\tilde{g}_{acbd}$$
$$\Omega_{aibj}^{B2} = \sum_{kl} t_{kl}^{ab}(\tilde{g}_{kilj} + \sum_{cd} t_{ij}^{cd}\tilde{g}_{kcld})$$
$$\Omega_{aibj}^{C2} = -\frac{1}{2}\sum_{ck} t_{kj}^{bc}\left(\tilde{g}_{kiac} - \frac{1}{2}\sum_{dl} t_{li}^{ad}\tilde{g}_{kdlc}\right) - \sum_{ck} t_{ki}^{bc}\left(\tilde{g}_{kjac} - \frac{1}{2}\sum_{dl} t_{lj}^{ad}\tilde{g}_{kdlc}\right)$$
$$\Omega_{aibj}^{D2} = \frac{1}{2}\sum_{ck} u_{jk}^{bc}\left(\tilde{L}_{aikc} + \frac{1}{2}\sum_{dl} u_{il}^{ad}\tilde{L}_{ldkc}\right)$$
$$\Omega_{aibj}^{E2} = \sum_{c} t_{ij}^{ac}({}^{I}\tilde{F}_{bc} - \sum_{dkl} u_{kl}^{bd}\tilde{g}_{ldkc}) - \sum_{k} t_{ik}^{ab}({}^{I}\tilde{F}_{kj} + \sum_{cdl} u_{lj}^{cd}\tilde{g}_{kdlc})$$

defined intermediates

$$u_{ij}^{ab} = 2t_{ij}^{ab} - t_{ji}^{ab}$$
$$\tilde{L}_{pqrs} = 2\tilde{g}_{pqrs} - \tilde{g}_{psrq}$$
$${}^{I}\tilde{F}_{rs} = \tilde{h}_{rs} + \sum_{i}(2\tilde{g}_{rsii} - \tilde{g}_{riis})$$

The CCSD amplitude equations in eqs 47 and 48 may be expressed in terms of the $\hat{T}_1$ transformed integrals as

$$\Omega_{ai}^{\mathrm{CCSD}} = \Omega_{ai}^{A1} + \Omega_{ai}^{B1} + \Omega_{ai}^{C1} + \Omega_{ai}^{D1} = 0 \qquad (55)$$

and

$$\Omega_{aibj}^{\mathrm{CCSD}} = \Omega_{aibj}^{A2} + \Omega_{aibj}^{B2} + P_{ij}^{ab}(\Omega_{aibj}^{C2} + \Omega_{aibj}^{D2} + \Omega_{aibj}^{E2}) = 0 \qquad (56)$$

where the explicit expressions for the individual terms in eqs 55 and 56 are given in Table 2, and where the P_{ij}^{ab} operator carries out the following permutation of pair indices:

$$P_{ij}^{ab}A_{ij}^{ab} = A_{ij}^{ab} + A_{ji}^{ba} \qquad (57)$$

The amplitude equations in eqs 55 and 56 constitute a set of nonlinear equations.

4.2. CCSD Amplitude Equations for Local HF Orbitals. We will now examine the locality of the CCSD amplitude equations in eqs 55 and 56. The CCSD amplitude equations are expressed in terms of $\hat{T}_1$ transformed integrals, which are defined in terms of the transformation matrices in eqs 53 and 54 involving the singles amplitudes. The singles amplitudes are small, and initially we ignore these amplitudes in the locality analysis. The CCSD amplitude equations in eqs 55 and 56 will thus be viewed as referencing conventional integrals in the local HF basis. Of course, when atomic fragment and atomic pair fragment energies are determined the $\hat{T}_1$ transformed integrals have to be considered. We elaborate on this point in Section 4.2.3 below.

A Møller−Plesset perturbation analysis shows that through first order in the fluctuation potential the singles equation vanishes and the doubles equation reduces to the MP2 amplitude equation in eq 17 (the first term in Ω^{A2} and the terms containing the Fock

Table 3. Locality Restrictions on the Summation Indices in the CCSD Equations[a]

contributions to singles residual	
$\Omega^{A1}_{ai} = 2\Sigma_{ckd} t^{cd}_{ki} \tilde{g}_{adkc} - \Sigma_{ckd} t^{cd}_{ik} \tilde{g}_{adkc}$	$a \in [\overline{P}]_2$, $c \in [\overline{R}]$, $d \in [\overline{P}]$, $k \in \underline{R}$; $a \in [\overline{P}]_4$, $c \in [\overline{P}]$, $d \in [\overline{P}]_3$, $k \in [\underline{P}]_2$
$\Omega^{B1}_{ai} = -2\Sigma_{ckl} t^{ac}_{kl} \tilde{g}_{kilc} + \Sigma_{ckl} t^{ac}_{lk} \tilde{g}_{kilc}$	$a \in [\overline{P}]_2$, $c \in [\overline{R}]$, $l \in \underline{R}$, $k \in [\underline{P}]$; $a \in [\overline{P}]_4$, $c \in [\overline{P}]_2$, $k \in [\underline{P}]$, $l \in [\underline{P}]_3$
$\Omega^{C1}_{ai} = 2\Sigma_{ck} {}^1\tilde{F}_{kc} t^{ac}_{ik} - \Sigma_{ck} {}^1\tilde{F}_{kc} t^{ac}_{ki}$	$a \in [\overline{P}]$, $c \in [\overline{R}]$, $k \in \underline{R}$; $a \in [\overline{P}]_3$, $c \in [\overline{P}]$, $k \in [\underline{P}]_2$
$\Omega^{D1}_{ai} = {}^1\tilde{F}_{ai}$	$a \in [\overline{P}]$
contributions to doubles residual	
$\Omega^{A2}_{aibj} = \tilde{g}_{aibj} + \Sigma_{cd} t^{cd}_{ij} \tilde{g}_{acbd}$	$a \in [\overline{P}]$, $b \in [\overline{Q}]$; $a \in [\overline{P}]_2$, $b \in [\overline{Q}]_2$, $c \in [\overline{P}]$, $d \in [\overline{Q}]$
$\Omega^{B2}_{aibj} = \Sigma_{kl} t^{ab}_{kl} \tilde{g}_{kilj} + \Sigma_{kl} t^{ab}_{kl} \Sigma_{cd} t^{cd}_{ij} \tilde{g}_{kcld}$	$a \in [\overline{P}]_2$, $b \in [\overline{Q}]_2$, $k \in [\underline{P}]$, $l \in [\underline{Q}]$; $a \in [\overline{P}]_3$, $b \in [\overline{Q}]_3$, $c \in [\overline{P}]$, $d \in [\overline{Q}]$, $k \in [\underline{P}]_2$, $l \in [\underline{Q}]_2$
$\Omega^{C2}_{aibj} = -(1/2)\Sigma_{ck} t^{bc}_{kj} \tilde{g}_{kiac}$ + $1/4\Sigma_{ck} t^{bc}_{kj} \Sigma_{dl} t^{ad}_{li} \tilde{g}_{kdlc} - \Sigma_{ck} t^{bc}_{ki} \tilde{g}_{kjac} + 1/2\Sigma_{ck} t^{bc}_{ki} \Sigma_{dl} t^{ad}_{ij} \tilde{g}_{kdlc}$	$a \in [\overline{Q}]_2$, $b \in [\overline{P}]_2$, $c \in [\overline{Q}]$, $k \in [\underline{P}]$; $a \in [\overline{Q}]_3$, $b \in [\overline{P}]_3$, $c \in [\overline{Q}]$, $d \in [\overline{P}]$, $k \in [\underline{P}]_2$, $l \in [\underline{Q}]_2$; $a \in [\overline{P}]_2$, $b \in [\overline{Q}]_2$, $c \in [\overline{P}]$, $k \in [\underline{Q}]$; $a \in [\overline{P}]_3$, $b \in [\overline{Q}]_3$, $c \in [\overline{P}]$, $d \in [\overline{Q}]$, $k \in [\underline{Q}]_2$, $l \in [\underline{P}]_2$
$\Omega^{D2}_{aibj} = 2\Sigma_{ck} t^{bc}_{jk} \tilde{g}_{aikc} - \Sigma_{ck} t^{bc}_{jk} \tilde{g}_{acki} - \Sigma_{ck} t^{bc}_{kj} \tilde{g}_{aikc} + 1/2\Sigma_{ck} t^{bc}_{kj} \tilde{g}_{acki} + 2\Sigma_{ck} t^{bc}_{jk} \Sigma_{dl} t^{ad}_{il} \tilde{g}_{ldkc}$ − $\Sigma_{ck} t^{bc}_{jk} \Sigma_{dl} t^{ad}_{il} \tilde{g}_{lckd} - \Sigma_{ck} t^{bc}_{jk} \Sigma_{dl} t^{ad}_{li} \tilde{g}_{ldkc} + 1/2\Sigma_{ck} t^{bc}_{jk} \Sigma_{dl} t^{ad}_{li} \tilde{g}_{lckd} - \Sigma_{ck} t^{bc}_{kj} \Sigma_{dl} t^{ad}_{il} \tilde{g}_{ldkc}$ + $1/2\Sigma_{ck} t^{bc}_{kj} \Sigma_{dl} t^{ad}_{il} g_{lckd} + 1/2\Sigma_{ck} t^{bc}_{kj} \Sigma_{dl} t^{ad}_{li} \tilde{g}_{ldkc} - 1/4\Sigma_{ck} t^{bc}_{kj} \Sigma_{dl} t^{ad}_{li} \tilde{g}_{lckd}$	$a \in [\overline{P}]$, $b \in [\overline{Q}]$, $c \in [\overline{R}]$, $k \in \underline{R}$; $a \in [\overline{P}]_3$, $b \in [\overline{Q}]$, $c \in [\overline{P}]_2$, $k \in [\underline{P}]$; $a \in [\overline{P}]$, $b \in [\overline{Q}]_3$, $c \in [\overline{Q}]$, $k \in [\underline{Q}]_2$; $a \in [\overline{Q}]_2$, $b \in [\overline{P}]_2$, $c \in [\overline{Q}]$, $k \in [\underline{P}]$; $a \in [\overline{P}]$, $b \in [\overline{Q}]$, $c \in [\overline{R}]$, $d \in [\overline{S}]$, $k \in \underline{R}$, $l \in \underline{S}$; $a \in [\overline{P}]$, $b \in [\overline{Q}]$, $c \in [\overline{R}]$, $k \in \underline{R}$, $l \in [\underline{R}]_2$, $d \in [\overline{R}]$; $a \in [\overline{P}]_3$, $b \in [\overline{Q}]$, $c \in [\overline{R}]$, $d \in [\overline{P}]$, $k \in \underline{R}$, $l \in [\underline{P}]_2$; $a \in [\overline{P}]_5$, $b \in [\overline{Q}]$, $c \in [\overline{P}]_3$, $d \in [\overline{P}]$, $k \in [\underline{P}]_2$, $l \in [\underline{P}]_4$; $a \in [\overline{P}]$, $b \in [\overline{Q}]_3$, $c \in [\overline{Q}]$, $d \in [\overline{R}]$, $k \in [\underline{Q}]_2$, $l \in \underline{R}$; $a \in [\overline{P}]$, $b \in [\overline{Q}]_5$, $c \in [\overline{Q}]$, $d \in [\overline{Q}]_3$, $k \in [\underline{Q}]_4$, $l \in [\underline{Q}]_2$; $a \in [\overline{P}]_3$, $b \in [\overline{Q}]_3$, $c \in [\overline{Q}]$,$d \in [\overline{P}]$,$k \in [\underline{Q}]_2$,$l \in [\underline{P}]_2$; $a \in [\overline{Q}]_3$, $b \in [\overline{P}]_3$, $c \in [\overline{Q}]$, $d \in [\overline{P}]$, $k \in [\underline{P}]_2$, $l \in [\underline{Q}]_2$
$\Omega^{E2}_{aibj} = +\Sigma_c t^{acl}_{ij} \tilde{F}_{bc} - \Sigma_k t^{ab}_{ik} {}^1\tilde{F}_{kj} - 2\Sigma_{ck} t^{ac}_{ij} \Sigma_{dl} t^{bd}_{kl} \tilde{g}_{ldkc} + \Sigma_{ck} t^{ac}_{ij} \Sigma_{dl} t^{bd}_{lk} \tilde{g}_{ldkc}$ − $2\Sigma_{cl} t^{ab}_{ik} \Sigma_{dk} t^{cd}_{lj} \tilde{g}_{kdlc} + \Sigma_{cl} t^{ab}_{ik} \Sigma_{dk} t^{cd}_{jl} \tilde{g}_{kdlc}$	$a \in [\overline{P}]$, $b \in [\overline{Q}]_2$, $c \in [\overline{Q}]$; $a \in [\overline{P}]$, $b \in [\overline{Q}]_2$, $k \in [\underline{Q}]$; $a \in [\overline{P}]$, $b \in [\overline{Q}]_3$, $c \in [\overline{Q}]$, $d \in [\overline{R}]$, $k \in [\underline{Q}]_2$, $l \in \underline{R}$; $a \in [\overline{P}]$, $b \in [\overline{Q}]_5$, $c \in [\overline{Q}]$, $d \in [\overline{Q}]_3$, $k \in [\underline{Q}]_2$, $l \in [\underline{Q}]_4$; $a \in [\overline{P}]$, $b \in [\overline{Q}]_3$, $c \in [\overline{R}]$, $d \in [\overline{Q}]$, $k \in [\overline{Q}]_2$, $l \in \underline{R}$; $a \in [\overline{P}]$, $b \in [\overline{Q}]_5$, $c \in [\overline{Q}]$, $d \in [\overline{Q}]_3$, $k \in [\underline{Q}]_4$, $l \in [\underline{Q}]_2$

[a] More precisely, the locality restrictions refer to the second CCSD iteration, but for clarity, the iteration number has been omitted from the amplitudes. In all terms it is assumed that $i \in \underline{P}$ and $j \in \underline{Q}$ and the remaining index restrictions have been derived based on this assumption. In some terms the k and l indices belong to general atomic sites R and S. However, R and S may also be restricted to indices local to P and Q without introducing significant errors, see the discussion in the text. In this way all terms may be evaluated in a local fragment orbital space. The locality constraints referencing different terms are separated by semicolons.

matrix in Ω^{E2}, see Table 2). Neglecting the second and higher order CCSD terms in the amplitude equations in eqs 55 and 56, the CCSD amplitude equations become identical to the MP2 amplitude equations, and these terms will therefore have the locality of the MP2 equations. In the following we examine how the second and higher order CCSD terms affect the locality of the MP2 amplitude equations, which was discussed in Section 3.

The first MP2 iteration determines the cluster amplitudes of the EOS: $i \in \underline{P}$, $a \in [\overline{P}]$, $j \in \underline{Q}$, $b \in [\overline{Q}]$. These amplitudes are used in the second iteration to determine the residual $R^{2,\mathrm{MP2}}_{aibj}(t^2)$ in eq 38 and to identify the FOI space. We now discuss the extensions of the FOI space that are required when the second and higher order terms in the CCSD amplitude equations are added to the MP2 residual $R^{2,\mathrm{MP2}}_{aibj}(t^2)$ in eq 38. In short, we examine the nonvanishing contributions to the residual, which originate from the second and higher order terms in $\Omega^{\mathrm{CCSD}}_{ai}(t^2)$ and $\Omega^{\mathrm{CCSD}}_{aibj}(t^2)$.

Using the locality restrictions for $t^{2,ab}_{ij}$ in eq 37 as a starting point, we may carry out a locality analysis for the CCSD equations in a similar manner as was done for the MP2 equations in Section 3. The results are summarized in Table 3, where for simplicity the iteration number for t^2 has been omitted from the amplitudes. From a locality perspective the residual terms in Table 3 can be divided into three types:

- Type 1: Terms where the locality is straightforwardly accounted for through the locality of the MP2 equations. These terms involve only P and possibly Q sites in Table 3.
- Type 2: Terms where the locality is accounted for through the locality in the MP2 equation but where the terms contain summation indices referencing contributions that are long ranged. These are terms involving other sites than P and Q in Table 3, i.e., sites R and possibly also S. The long-ranged decay, however, does not follow the behavior of a Coulombic potential, which decays with inverse distance between the atomic sites P and R and between atomic sites Q and R. Rather it follows an inverse cubed distance behavior as we discuss in more detail below.
- Type 3: Terms involving the $\hat{T}_1$ transformed Fock matrix $\tilde{\mathbf{F}}$.

4.2.1. Type 1 Terms. These terms include most of the second and higher order terms in eqs 55 and 56. An example is the second term in Ω^{A2} ($\Sigma_{cd} t^{cd}_{ij} \tilde{g}_{acbd}$) which may be expressed as

$$I^{ab}_{ij} = \sum_{c \in [\overline{P}], d \in [\overline{Q}]} t^{cd}_{ij} \tilde{g}_{acbd}; \quad i \in \underline{P}, \quad j \in \underline{Q}, \quad a \in [\overline{P}]_2, \quad b \in [\overline{Q}]_2 \tag{58}$$

To obtain eq 58 the summation indices have been restricted to $c \in [\overline{P}]$, $d \in [\overline{Q}]$ using eq 37 and a nonvanishing integral $\tilde{g}_{acbd}$ then implies that $a \in [\overline{P}]_2$ and $b \in [\overline{Q}]_2$. The locality analysis for the MP2 amplitude equations ensures that this term is properly treated.

For some terms of type 1 the coupling to neighboring sites is more extensive than for MP2, e.g., for the second term in B2 where $a \in [\overline{P}]_3$ and $b \in [\overline{Q}]_3$. However for this term the more extensive coupling is due to a term that is of third order in Møller−Plesset perturbation theory and therefore has a very small effect on the solution of the amplitude equations in the

EOS. If it does have a nonvanishing effect, then it is taken into account through our way of determining orbital buffer spaces.

4.2.2. Type 2 Terms. An example of these terms is the first term in Ω^{D2} $(1/2\Sigma_{ck}u_{jk}^{bc}\tilde{L}_{aikc})$ which has a contribution:

$$J_{ij}^{ab} = \sum_{R} \sum_{c \in [\overline{R}], k \in \underline{R}} t_{jk}^{bc}\tilde{g}_{aikc};\ \ i \in \underline{P},\ \ j \in \underline{Q},\ \ a \in [\overline{P}], b \in [\overline{Q}] \tag{59}$$

where the c summation has been restricted because $\tilde{g}_{aikc}$ has to be nonvanishing. In principle, the R summation runs over all atomic sites, however, the integrals $\tilde{g}_{aikc}$ and the amplitudes t_{jk}^{bc} both have a cubic decay with respect to the distance between the atomic sites P and R and between the atomic sites Q and R, respectively. The cubic decay of integrals and amplitudes is due to the fact that when a multipole expansion of the integral $\tilde{g}_{aikc}$ is carried out, the monopole integral contribution vanishes.[34] This happens because the charge distributions carry no charge as the occupied, and virtual orbitals are orthogonal. Furthermore, we recall that this term is second order in the fluctuation potential and therefore not one of the main contributors to the total residual. For these reasons it is well-justified to restrict the R summation in eq 59 to include only atoms local to P and Q.

For other doubles amplitude terms in Table 3 (e.g., $2\Sigma_{ck}t_{jk}^{bc}\Sigma_{dl}t_{il}^{ad}\tilde{g}_{ldkc}$ in Ω^{D2}), the summations occur over two atomic sites R and S. The cubic decay of amplitudes and integrals and the fact that these terms are only third-order in the fluctuation potential justify that they can be evaluated within a local fragment orbital space.

For the singles amplitude contributions (e.g., $2\Sigma_{ckd}t_{ki}^{cd}\tilde{g}_{adkc}$ in Ω^{A1}) the integral only has a quadratic inverse decay because the $\phi_a\phi_d$ distribution does carry a charge. The t_{ki}^{cd} amplitudes, however, have cubic decay with respect to the distance between the charge distributions $\phi_k\phi_c$ and $\phi_i\phi_d$ due to orthogonality.

4.2.3. Type 3 Terms. When carrying out the fragment calculations the $\hat{T}_1$-transformed Fock matrix ${}^{I}\tilde{\mathbf{F}}$:

$${}^{I}\tilde{F}_{rs} = \tilde{h}_{rs} + \sum_{i}(2\tilde{g}_{rsii} - \tilde{g}_{riis}) \tag{60}$$

has long-range contributions as a result of the $\hat{T}_1$ transformation, see eqs 51 and 52.

When evaluating atomic fragment and atomic pair fragment energies we may introduce the following approximation to the $\hat{T}_1$-transformed Fock matrix:

$${}^{I}\tilde{F}_{pq} \rightarrow {}^{I}\tilde{F}_{pq}^{\mathcal{A}} - {}^{I}F_{pq}^{\mathcal{A}} + {}^{I}F_{pq} \tag{61}$$

where ${}^{I}F_{pq}^{\mathcal{A}}$ is a nontransformed Fock matrix element referencing only the orbitals of the fragment orbital space, and ${}^{I}\tilde{F}_{pq}^{\mathcal{A}}$ is defined in the same way but with $\hat{T}_1$-transformed integrals. Within this approximation we have thus only included the effect of the $\hat{T}_1$ transformation inside the fragment, whereas the polarization from orbitals outside the fragment orbital space has been neglected.

The solution to this problem is to save all singles amplitudes when carrying out the atomic fragment and the atomic pair fragment calculations for all fragments in the molecule. In this first calculation we use the Fock matrix in eq 61. After all fragment calculations have been carried out we may use the stored singles amplitudes to construct the fully $\hat{T}_1$-transformed Fock matrix ${}^{I}\tilde{\mathbf{F}}$, which include the polarization of all orbitals. A new calculation may then be carried out using the fully transformed Fock matrix instead of eq 61. In this second calculation the long-ranged polarization effects described by the singles amplitudes are treated to a high precision. An even more elaborate treatment of the singles polarization effect is to use eq 61, where ${}^{I}F_{pq}$ and ${}^{I}F_{pq}^{\mathcal{A}}$ are replaced by the one-index transformed Fock matrix with t-amplitudes from the first calculation.

4.3. Summary of CCSD Locality Analysis. Even though the CCSD locality analysis is much more cumbersome than the corresponding analysis for MP2, our analysis shows that it is indeed possible to carry out a full CCSD calculation in terms of small independent fragment calculations. In other words, the analysis substantiates from a theoretical point of view that the CCSD fragment calculations may be carried out in orbital spaces of limited size. More precisely, all terms in the CCSD residual are local in the sense that the effects on the t_{ij}^{ab} amplitudes required for pair fragment PQ (i.e., $i \in \underline{P}$ and $j \in \underline{Q}$) rapidly decrease as the summation indices in the individual residual contributions move away from atomic sites P and Q. One should also keep in mind that the first-order residual terms (MP2) are the dominating ones, and the main effect of the second and higher order CCSD terms is to relax the amplitudes. The sizes of the EOSs will be similar for MP2 and CCSD, because they are basically obtained by requiring nonzero overlap distributions for the integrals entering the fragment energies, see eqs 8 and 9. However, the CCSD fragments will in general require larger buffer orbital spaces than MP2 fragments due to a more extensive coupling of amplitudes in CCSD.

The analysis above is carried out for an atomic pair fragment ($i \in \underline{P}$ and $j \in \underline{Q}$). It follows that for an atomic fragment (corresponding to $P = Q$) the effects of the CCSD residual contributions diminish with increasing distance between the atomic site P and the orbital summation indices in the CCSD residual. In an atomic fragment calculation, we gradually extend the orbital fragment sizes to include orbitals located still further away from atomic site P (to be detailed in Section 6.2). Since all contributions fall of rapidly with increasing distance to the P center, each step of the fragment expansion procedure will include orbitals with still smaller effects on the fragment energy. This orbital expansion procedure can therefore be repeated until the atomic fragment energy has been converged to a predefined threshold. Of course we do not know, for a given orbital space expansion, which of the individual CCSD residual terms have the largest effect on the atomic fragment energy, but the fragment expansion procedure ensures that the orbital spaces are large enough that only the truly small terms are neglected. If the optimal orbital spaces for atomic fragments P and Q have been determined, the analysis above then substantiates that the corresponding pair fragment PQ can be formed from the union of orbital spaces for P and Q.

5. COMPUTATIONAL SCALING IN DEC CC CALCULATIONS

Within the DEC model a CC calculation is carried out in terms of atomic fragment and atomic pair fragment calculations. The locality of the CC calculation is determined by the atomic fragment calculations. The sizes of the amplitude orbital spaces are determined in a black box manner in a way that ensures that the atomic fragment energies are calculated to a preset threshold. The orbital spaces employed for evaluating atomic pair fragment energies are determined as unions of the orbital spaces for the two atomic fragments involved.

Since the major task of a CC calculation is to describe short-ranged phenomena associated with Coulomb holes in the wave function and the dispersion effects, the fragment sizes are to a large extent system independent. The number of atomic fragments scales as the number of atoms in the molecular system, and the calculation of atomic fragment energies is therefore linearly scaling. The number of atomic pair fragment energies in eq 5 has a quadratic scaling with system size, but this scaling is reduced to linear for large systems

because the pair fragment calculations only need to be carried out for pair atomic distances, where the dispersion forces are non-negligible.

The DEC scheme is embarrassingly parallelizable. The evaluation of the atomic fragment energies may first be done independently; second the atomic pair fragment energies that are needed may also be calculated independently. Provided that a sufficient number of processors are available, the wall time for a parallelized correlated wave function calculation is therefore the sum of a single atomic fragment calculation and a single atomic pair fragment energy calculation, not considering the time for the initial HF calculation, whose efficient evaluation has been discussed previously, and may be done using a linear-scaling algorithm.[35−38]

6. DETAILS ABOUT THE DEC MODEL

In Section 6.1 we discuss how to improve the locality of the molecular orbitals (MOs) used in a DEC fragment calculation. In Section 6.2 we discuss how the sizes of the EOS and AOS are determined. The fragment calculations are carried out as standard MP2 and CCSD calculations, as we detail in Section 6.3. In Section 6.4 we discuss how errors similar to basis set superposition errors (BSSE) may be avoided when calculating pair interaction energies ΔE_{PQ}. Finally, in Section 6.5 we summarize the various spaces employed in a DEC fragment calculation.

6.1. Defining Locality of MOs and Atomic Fragment Extents. Even though the LCM HF orbitals[30] are local, the orthonormal character of this basis causes small but nonvanishing coefficients on atoms located some distance away from the atomic site where the HF orbital was assigned. In this section we discuss how this tail region may be treated without considering explicitly the atomic sites at which the small expansion coefficients are situated. This development will be used to reduce the number of atomic sites where the atomic orbital integrals have to be evaluated.

A normalized HF orbital associated with atomic site P:

$$\varphi_r^P = \sum_\mu \chi_\mu c_{\mu r}^P \tag{62}$$

may be approximated in the following way:

$$\tilde{\varphi}_r^P = \sum_{\tilde{\mu}} \chi_{\tilde{\mu}} \tilde{c}_{\tilde{\mu} r}^P \tag{63}$$

where the $\tilde{\mu}$-summation is restricted to atomic sites, which in some sense are neighboring the atomic site P (to be detailed below). For example, the tail region of ϕ_r^P may be excluded. The expansion coefficients of $\tilde{\phi}_r^P$ may be determined from a least-squares fit of

$$f(\tilde{c}^P) = \| \tilde{\varphi}_r^P - \varphi_r^P \| \tag{64}$$

giving the expansion coefficients

$$\tilde{c}_{\tilde{\mu} r}^P = \sum_{\tilde{\nu}\eta} \tilde{S}_{\tilde{\mu}\tilde{\nu}}^{-1} S_{\tilde{\nu}\eta} c_{\eta r}^P \tag{65}$$

where the dimensions of the overlap matrices are defined by the restrictions that are imposed on the AO indices:

$$\tilde{S}_{\tilde{\mu}\tilde{\nu}} = \langle \chi_{\tilde{\mu}} | \chi_{\tilde{\nu}} \rangle \tag{66}$$

$$S_{\tilde{\nu}\eta} = \langle \chi_{\tilde{\nu}} | \chi_\eta \rangle \tag{67}$$

We note that an equation similar to eq 65 was also used by Usvyat and Schütz[39] for obtaining localized Wannier functions. We now discuss how the $\tilde{\mu}$-summation in eq 63 may be restricted to exclude the tail region of atomic fragments. An HF orbital ϕ_r^P was assigned to atomic site P because its largest Mulliken charge was situated on atomic site P. For ϕ_r^P all nonvanishing Mulliken charges may be determined and arranged in order of decreasing size to prioritize the importance of the atomic sites. We may then restrict $\tilde{\mu}$ to the atomic sites which have a Mulliken charge larger than a given threshold and determine the expansion coefficients $\tilde{c}_{\tilde{\mu} r}^P$ from eq 65. Using $(1 - \langle \tilde{\phi}_r^P | \tilde{\phi}_r^P \rangle)$ as a measure of the quality of the least-squares fit, we identify the largest Mulliken charge threshold for which

$$1 - \langle \tilde{\varphi}_r^P | \tilde{\varphi}_r^P \rangle = 1 - \sum_{\tilde{\mu}\tilde{\nu}} \tilde{c}_{\tilde{\mu} r}^P \tilde{S}_{\tilde{\mu}\tilde{\nu}} \tilde{c}_{\tilde{\nu} r}^P < \delta \tag{68}$$

where δ is a small prefixed number. The atomic sites defined by the Mulliken charge threshold determine the orbital extent of $\tilde{\phi}_r^P$. The union of extents for all HF orbitals assigned to atomic site P is denoted the atomic extent $\{P\}$. We note that a screening of atomic centers in accordance with eq 68 was used by Boughton and Pulay[32] as a completeness criteria for the assignment of excitation spaces for the occupied HF orbitals.

When solving the amplitude equations for the atomic fragment P calculation we use the MOs in the AOS $\mathscr{A}_P$ in eq 45. The MOs have expansion coefficients on atomic sites outside the $\mathscr{A}_P$ space. The atomic sites, where atomic integrals have to be evaluated to ensure that the MO integrals in $\mathscr{A}_P$ are properly evaluated, consist of the union of atomic extents for all atoms in $\mathscr{A}_P$. We denote this space the atomic fragment extent $\{\mathscr{A}_P\}$. If for example we consider the case where $\mathscr{A}_P$ contains the atomic sites L, M, N, P, Q, and R, then

$$\{\mathscr{A}_P\} = \{L\} \cup \{M\} \cup \{N\} \cup \{P\} \cup \{Q\} \cup \{R\} \tag{69}$$

We have now defined the amplitude orbital space $\mathscr{A}_P$, where the amplitude calculation is carried out, and the atomic centers in terms of which the MOs are expanded $\{\mathscr{A}_P\}$. The MOs that are used in an atomic fragment calculation are then obtained from an expansion of the form

$$\tilde{\varphi}_r^X = \sum_{\tilde{\mu} \in \{\mathscr{A}_P\}} \chi_{\tilde{\mu}} \tilde{c}_{\tilde{\mu} r}^X; \qquad (X \in \mathscr{A}_P) \tag{70}$$

where the summation is restricted to $\{\mathscr{A}_P\}$. The expansion coefficients in eq 70 may be determined from eq 65, giving the best uniform description of the MOs in $\mathscr{A}_P$ confined to the space $\{\mathscr{A}_P\}$. The quality of the least-squares fit, which determine the MOs, depends on the parameter δ in eq 68. Reducing the size of δ will lead to an extension of $\{\mathscr{A}_P\}$ and thus a better least-squares fit.

We may also require that the approximated orbitals in eq 70 are normalized

$$\langle \tilde{\varphi}_r^X | \tilde{\varphi}_r^X \rangle = \sum_{\tilde{\mu}\tilde{\nu}} \tilde{c}_{\tilde{\mu} r}^X \tilde{S}_{\tilde{\mu}\tilde{\nu}} \tilde{c}_{\tilde{\nu} r}^X = 1 \tag{71}$$

by multiplying $\tilde{\phi}_r^X$ by an appropriate constant such that eq 71 is satisfied.

We have now established the locality of an atomic fragment calculation in terms of the spaces $\mathscr{A}_P$ and $\{\mathscr{A}_P\}$ and identified the MOs that should be used in the amplitude calculations. For clarity let us summarize the main steps for constructing the approximate MOs $\tilde{\phi}$:

- Choose a δ value which is a measure of how much the approximated orbitals $\tilde{\phi}$ deviate from the original orbitals ϕ.
- For each approximated orbital add atoms based on their Mulliken charge until eq 68 is satisfied. This newly generated list of atoms constitutes the orbital extent for each orbital.
- For each atom P construct the atomic extent as the union of orbital extents for all orbitals assigned to that atom.

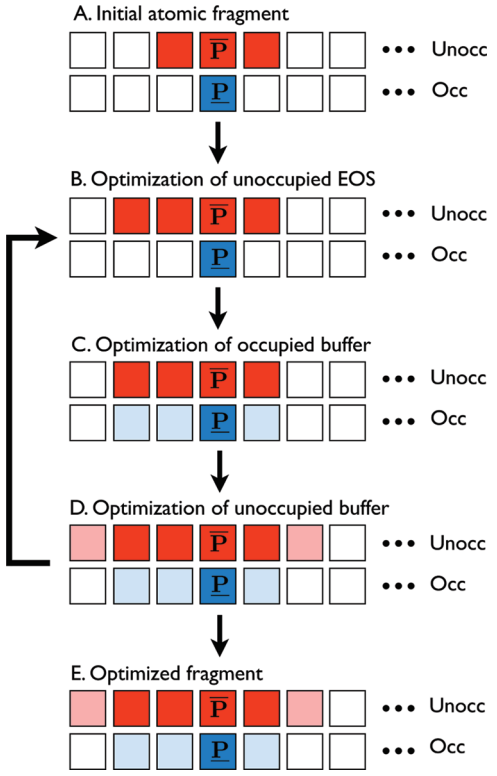

Figure 3. Main steps in the fragment optimization procedure. (A) Initially a starting guess is made for the size of the molecular fragment for evaluating E_P. The unoccupied energy orbital space (red) is then determined (B) followed by an optimization of the occupied (light blue) and unoccupied (pink) buffer spaces (C and D). It is then checked that the spaces are still optimal going through the steps in B−D again until the sizes of all spaces remain unchanged to yield the optimized fragment (E).

- Construct the atomic fragment extent $\{\mathscr{A}_P\}$ as the union of atomic extents for the atomic sites in $\mathscr{A}_P$.
- Determine the local MOs using eq 70, where the MO coefficients are determined from eq 65. Subsequently the MOs may be normalized by enforcing eq 71.

The above development may be viewed as an effective integral screening technique which allows atomic sites that reference small molecular orbital expansion coefficients to be removed from the atomic integral evaluation. The screening depends on the single parameter δ in eq 68.

6.2. Optimization of Fragment Orbital Spaces. An atomic fragment calculation for fragment P is carried out using the atomic fragment AOS $\mathscr{A}_P$ in Figure 2, and the atomic fragment energy is calculated from the amplitudes referencing the atomic fragment EOS $\mathscr{E}_P$, i.e., $\underline{P}$ and $[\overline{P}]$ for the occupied and unoccupied spaces, respectively. The orbital spaces $[\overline{P}]$, $\mathscr{B}_{\underline{P}}$, and $\mathscr{B}_{\overline{P}}$ are determined from a sequence of calculations, where the sizes of these spaces are gradually increased until the atomic fragment energy E_P in eq 8 is converged to a preset energy threshold. In Figure 3A−D we have illustrated how the unoccupied EOS $[\overline{P}]$ (red) and the buffer spaces $\mathscr{B}_{\underline{P}}$ (light blue) and $\mathscr{B}_{\overline{P}}$ (pink) may gradually be increased to ensure that E_P is determined to a preset energy tolerance, the fragment optimization threshold (FOT). When the change in the atomic fragment energy is larger than the FOT, the step is accepted, otherwise it is rejected. Continuing in this manner the cycle in Figure 3B−D is repeated until we have carried through one full optimization round where the sizes of the unoccupied EOS, occupied buffer, and unoccupied buffer spaces all remain unchanged.

In practice the sizes of the unoccupied EOS, occupied buffer, and unoccupied buffer spaces are defined by three orbital space radii. For example, if the distance between atomic site P and an atomic site R is smaller than the unoccupied EOS radius for P, then the unoccupied orbitals assigned to R are included in the unoccupied EOS $[\overline{P}]$. The increment of orbital spaces may thus be controlled by adding a fixed orbital space step size parameter to the existing orbital space radii.

The atomic pair fragment orbital space for evaluating E_{PQ} is formed as the union of the atomic fragment orbital spaces for evaluating E_P and E_Q. The locality analysis in Sections 3.2 and 4.2 substantiates that it is not necessary to carry out additional fragment size optimizations for the combined PQ space, as the locality is defined by the atomic fragment calculations on P and Q. The numerical results presented in Section 7 also support this.

The fragment optimization procedure discussed above applies both to MP2 and CCSD fragment calculations. In a CCSD calculation we may take advantage of the fact that the locality of the CCSD amplitude equations is similar to the one for MP2 equations, except for a more extensive coupling among the amplitudes. Therefore the optimized MP2 fragment spaces serve as good starting guesses for the optimal CCSD fragment orbital spaces.

6.3. CC Calculations on Atomic Fragments and Atomic Pair Fragments. The MP2 and CCSD atomic fragment and atomic pair fragment calculations are carried out as standard MP2 and CCSD calculations using the local HF orbitals and the atomic fragment or atomic pair fragment orbital spaces. The MP2 amplitude equations constitute a set of linear equations with a symmetric matrix and are solved using the conjugate residual with optimal trial vectors (CROP) algorithm, where only the last three trial vectors need to be stored to maintain the information content of all previous trial vectors (see ref 40). For CCSD we solve nonlinear amplitude equations using the CROP algorithm. The nonlinearity of the CCSD amplitude equations generally requires more than three trial vectors to be stored.[40]

6.4. Avoiding Wave Function Superposition Problems in Pair Energy Calculations. The calculation of pair interaction energies requires special attention because the equations presented in Section 2.2 may be subject to errors which are similar to BSSE. We now demonstrate how this problem may easily be circumvented.

The nontruncated expression for the pair interaction energy is given in eq 6:

$$\begin{aligned}\Delta E_{PQ} &= E_{PQ} - E_P - E_Q \\ &= \sum_{ij \in \underline{P} \cup \underline{Q}}^{ab} (t_{ij}^{ab} + t_i^a t_j^b)(2g_{iajb} - g_{ibja}) \\ &\quad - \sum_{ij \in \underline{P}}^{ab} (t_{ij}^{ab} + t_i^a t_j^b)(2g_{iajb} - g_{ibja}) \\ &\quad - \sum_{ij \in \underline{Q}}^{ab} (t_{ij}^{ab} + t_i^a t_j^b)(2g_{iajb} - g_{ibja}) \end{aligned} \quad (72)$$

where we have inserted the expressions for the atomic fragment and the atomic pair fragment energies in eqs 3 and 4. No approximations have been introduced in eq 72 since the virtual summation indices run over the full molecular system.

In DEC fragment calculations we may evaluate ΔE_{PQ} in two different ways: (i) the simple approach or (ii) the counterpoise-corrected approach.

For case (i) we simply insert the approximate expressions for E_P, E_Q, and E_{PQ} in eqs 8 and 9 into eq 6:

$$\begin{aligned}\Delta E_{PQ} &= E_{PQ} - E_P - E_Q \\ &= \sum_{\substack{ij \in \underline{P} \cup \underline{Q} \\ ab \in [\overline{P}] \cup [\overline{Q}]}} (t^{ab}_{PQ,ij} + t^{a}_{PQ,i} t^{b}_{PQ,j})(2g_{iajb} - g_{ibja}) \\ &\quad - \sum_{\substack{ij \in \underline{P} \\ ab \in [\overline{P}]}} (t^{ab}_{P,ij} + t^{a}_{P,i} t^{b}_{P,j})(2g_{iajb} - g_{ibja}) \\ &\quad - \sum_{\substack{ij \in \underline{Q} \\ ab \in [\overline{Q}]}} (t^{ab}_{Q,ij} + t^{a}_{Q,i} t^{b}_{Q,j})(2g_{iajb} - g_{ibja}) \end{aligned} \quad (73)$$

where we have added P, Q, and PQ subscripts to emphasize which fragment space is used for determining the amplitudes. The problem with eq 73 is that the amplitudes used for determining E_P, E_Q, and E_{PQ} are obtained from fragment calculations in different AOSs , i.e., $\mathcal{A}_P$, $\mathcal{A}_Q$ and $\mathcal{A}_{PQ} = \mathcal{A}_P \cup \mathcal{A}_Q$. The orbital space employed for the atomic pair fragment PQ calculation is thus larger than the orbital spaces used for the P and Q atomic fragment calculations. This means that the amplitudes used to determine E_{PQ} are determined slightly more accurately than the amplitudes used to determine E_P and E_Q, and therefore BSSE-like errors will be introduced when the pair interaction energy is calculated according to eq 73.

For case (ii), to avoid BSSE-like errors, we may calculate a counterpoise-corrected pair interaction energy by using only amplitudes from the pair fragment calculation and letting all virtual summations run over the unoccupied EOS in the pair fragment:

$$\begin{aligned}\Delta E_{PQ} &= \sum_{\substack{ij \in \underline{P} \cup \underline{Q} \\ ab \in [\overline{P}] \cup [\overline{Q}]}} (t^{ab}_{PQ,ij} + t^{a}_{PQ,i} t^{b}_{PQ,j})(2g_{iajb} - g_{ibja}) \\ &\quad - \sum_{\substack{ij \in \underline{P} \\ ab \in [\overline{P}] \cup [\overline{Q}]}} (t^{ab}_{PQ,ij} + t^{a}_{PQ,i} t^{b}_{PQ,j})(2g_{iajb} - g_{ibja}) \\ &\quad - \sum_{\substack{ij \in \underline{Q} \\ ab \in [\overline{P}] \cup [\overline{Q}]}} (t^{ab}_{PQ,ij} + t^{a}_{PQ,i} t^{b}_{PQ,j})(2g_{iajb} - g_{ibja}) \end{aligned} \quad (74)$$

We thereby determine ΔE_{PQ} in line with a counterpoise-corrected interaction energy.

6.5. Overview of the Spaces Employed in a DEC Calculation. It is now in place to summarize the various spaces employed in the atomic fragment and atomic pair fragment calculations.

In Table 1 we summarize the notation used for the various energies and orbital spaces used in a DEC calculation, while Figure 2 gives an illustrative overview of the orbital spaces used in an atomic fragment calculation. The amplitude equation for atomic fragment P is solved for MOs assigned to atoms in the AOS $\mathcal{A}_P$ in eq 45, and the fragment energy E_P is calculated using only the EOS orbitals, see eq 8. $\mathcal{E}_P$ and $\mathcal{A}_P$ are optimized during the calculation as discussed in Section 6.2. The MOs in atomic fragment P have nonvanishing expansion coefficients only for the atomic sites in the $\{\mathcal{A}_P\}$ space. Thus, when solving the amplitude equations in the $\mathcal{A}_P$ space, we need two-electron integrals in the AO basis for atomic sites in the $\{\mathcal{A}_P\}$ space to properly describe the necessary MO integrals.

For the atomic pair fragment calculations we employ the union of spaces from the atomic fragment calculations as discussed in Section 3.2. The amplitude equations are solved using the $\mathcal{A}_P \cup \mathcal{A}_Q$ AOS space, and the pair interaction energy ΔE_{PQ} is calculated in the $\mathcal{E}_P \cup \mathcal{E}_Q$ EOS space as described in Section 6.4. The MOs are expanded in terms of atoms in the $\{\mathcal{A}_P\} \cup \{\mathcal{A}_Q\}$ space, and two-electron integrals in the AO basis are therefore calculated in that space.

7. ILLUSTRATIVE RESULTS

In Section 7.1 we present calculations demonstrating that the total MP2 and CCSD correlation energies may be determined from DEC fragment calculations with control of the errors introduced compared to a full molecular calculation. In Section 7.2 we show that the fragment sizes and the relative energy errors in DEC-MP2 calculations are independent of the size of the molecular system.

7.1. DEC-MP2 and DEC-CCSD Calculations on $C_{14}H_2$. Using the $C_{14}H_2$ molecule as a test system we now compare DEC-MP2 and DEC-CCSD calculations to full molecular MP2 and CCSD calculations. We first discuss the assignment of orbitals in Section 7.1.1. In Section 7.1.2 we discuss the energy errors compared to a full molecular calculation as a function of the FOT. We demonstrate the importance of using counterpoise-corrected pair interaction energies in Section 7.1.3, and in Section 7.1.4 we discuss the sizes of the orbital spaces employed. Finally, in Section 7.1.5 we show that the total correlation energy is very insensitive to the choice of δ parameter. For the calculations in Sections 7.1.2−7.1.4 we have used $\delta = 0.01$ and imposed the normalization condition in eq 71. Furthermore, we have used orbital space step size parameters of 3.0 au for all orbital spaces (see Section 6.2),

7.1.1. Assignment of Orbitals. The number of occupied and virtual orbitals assigned to each atomic site in the $C_{14}H_2$ molecule are given in Figure 1B. The virtual orbitals are evenly distributed among the different atomic sites, while the distribution is less even for the occupied orbitals. The hydrogen atoms in general have no occupied orbitals assigned, whereas the carbon atoms have between two and four occupied orbitals assigned. This orbital partitioning may seem counterintuitive as we are considering a highly symmetric molecule. It is, however, not surprising, because the LCM orbitals do not reflect the molecular point group symmetry and are simply assigned to atoms based on the largest Mulliken charge. Therefore, if for a given molecular orbital two atoms have roughly the same Mulliken charge, then

the assignment of that orbital is quite arbitrary. However, the precise assignment does not affect the total correlation energy. The DEC algorithm automatically adjusts the size of the virtual orbital space for an uneven assignment of occupied orbitals through the way orbital spaces are selected (see Figure 3). For example, in Figure 1B, a carbon atom with four occupied orbitals assigned will, in general, require a larger virtual orbital space than a carbon atom with only two occupied orbitals assigned.

7.1.2. Energy Errors. In Table 4 we have listed the total energy errors (fourth column) compared to a full molecular calculation for various FOTs for MP2 and CCSD. It is seen that in general the errors decrease by an order of magnitude when the FOT is lowered by an order of magnitude, even though some deviance from this general result is observed for CCSD (to be explained below).

The second and third columns in Table 4 contain the sums of the single atomic fragment and the atomic pair interaction energy errors compared to a full space calculation. We henceforth denote $\Sigma_P E_P$ as the total single energy and $\Sigma_{P>Q}\Delta E_{PQ}$ as the total pair interaction energy. For both MP2 and CCSD, the total single energy and the total pair interaction energy errors systematically decrease when tightening the FOT. Furthermore, total single and total pair interaction energy errors are of similar magnitude, substantiating that pair fragments may be determined as unions of atomic fragments without carrying out additional fragment optimizations.

Table 4. Energy Errors [au] Compared to a Full Molecular Calculation for Single, Pair, And Total Correlation Energies as a Function of Fragment Optimization Thresholds[a]

threshold	$\Delta(\Sigma_P E_P)$	$\Delta(\Sigma_{P>Q}\Delta E_{PQ})$	ΔE_{corr}	% of E_{corr}
		MP2		
10^{-3}	0.013063	0.016875	0.029938	98.278
10^{-4}	0.001601	0.001775	0.003377	99.806
10^{-5}	0.000121	0.000139	0.000260	99.985
10^{-6}	0.000023	0.000014	0.000037	99.998
		CCSD		
10^{-3}	−0.000811	0.000369	−0.000442	100.025
10^{-4}	−0.000255	0.000067	−0.000189	100.011
10^{-5}	−0.000059	0.000046	−0.000013	100.001
10^{-6}	−0.000021	0.000006	−0.000015	100.001

[a] The calculations were carried out on the $C_{14}H_2$ molecule at the MP2 and CCSD levels of theory using the cc-pVDZ basis.

In general, all MP2 errors are positive, i.e., the fragment energies calculated using the DEC scheme are larger than fragment energies calculated in the total orbital space. Therefore the total energy error ΔE_{corr} compared to a full molecular calculation also decreases systematically when the FOT is lowered. In contrast, for CCSD the total single energy errors are negative, whereas the total pair interaction energy errors are positive. This leads to cancellation of errors for CCSD, and therefore the total energy errors for CCSD are smaller than those for MP2. This cancellation of errors is also the reason why the error (accidentally) is smaller when the FOT is 10^{-5} than 10^{-6} for the CCSD case, although the total single and total pair interaction energy errors both decrease when lowering the threshold. A closer look at the individual atomic fragment energy errors reveals that these are always positive for MP2, whereas their signs vary for CCSD. This will in general lead to cancellation of errors for DEC-CCSD compared to DEC-MP2. Thus, DEC-MP2 shows the "worst case scenario", where all individual fragment errors are added, whereas DEC-CCSD in general give smaller errors due to cancellation of errors. However, if the amplitudes are used to evaluate for example molecular gradients, errors of similar size will be obtained for MP2 and CCSD, as the cluster amplitudes are of similar quality.

In summary, it is a crucial feature of the DEC model that the full molecular correlation energy may be determined to any desired accuracy simply by choosing the appropriate FOT. It should be noted though that for CCSD, the error may be smaller than expected due to fortuitous cancellation of fragment energy errors.

7.1.3. Counterpoise Corrections. In the calculations presented so far we have calculated the pair interaction energies using the counterpoise-corrected expression in eq 74. In Figure 4 we have plotted the total pair interaction energy error against the FOT using the counterpoise-corrected expression (red solid curve) and using the simple expression in eq 73 (green dashed curve), which is prone to BSSE-like errors. The counterpoise-corrected strategy is clearly superior to the simple approach, and we therefore use counterpoise-corrected pair interaction energies in the calculations presented in this paper.

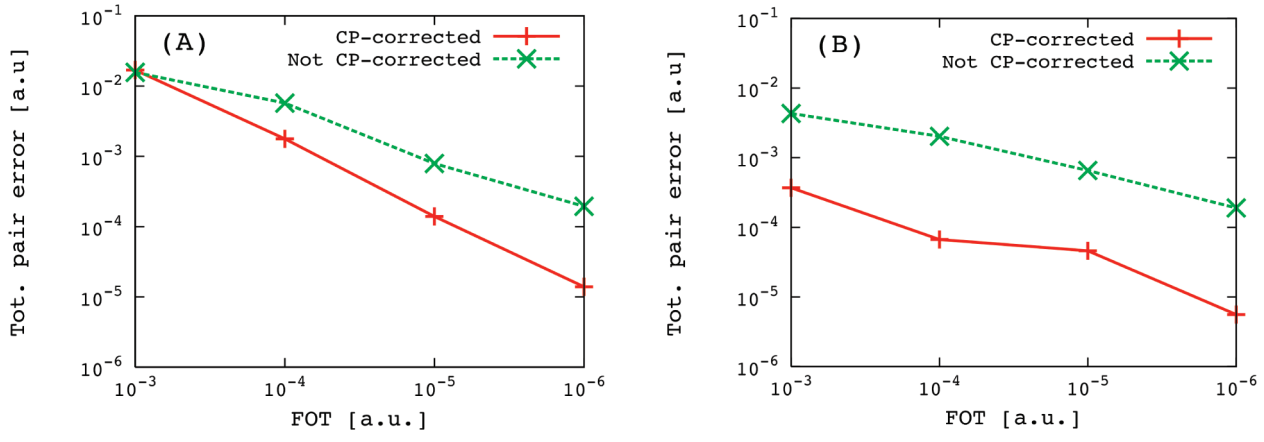

Figure 4. Errors of the total pair interaction energy $\Sigma_{P>Q}\Delta E_{PQ}$ compared to a full molecular calculation as a function of the FOT for MP2/cc-pVDZ (A) and CCSD/cc-pVDZ (B) calculations on the $C_{14}H_2$ molecule. Red solid curve: counterpoise-corrected (CP) pair interaction energies, see eq 74. Green dashed curve: pair interaction energies obtained by simple subtraction of atomic fragment energies from pair energies, see eq 73.

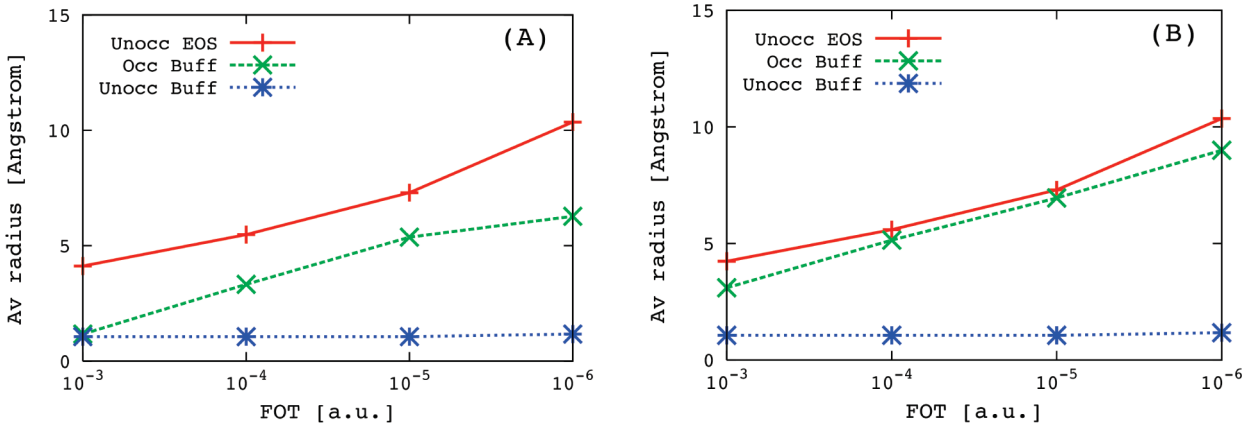

Figure 5. Average orbital space radii for DEC-MP2/cc-pVDZ (A) and DEC-CCSD/cc-pVDZ (B) calculations on the $C_{14}H_2$ molecule. Red solid curve: unoccupied energy orbital space. Green dashed curve: occupied buffer space. Blue dotted cuve: buffer unoccupied space.

7.1.4. Sizes of Orbital Spaces. In Figure 5 we have plotted the average orbital space radii for various FOTs. Recall that the orbital space radii define the number of MOs in the atomic fragments, see the discussion in Section 6.2. To avoid confusion we note that the unoccupied buffer space radius refers only to the buffer space, i.e., the total unoccupied space radius is the sum of the unoccupied EOS radius and the unoccupied buffer radius. As expected the orbital space radii in Figure 5 increases when the FOT is lowered, i.e., the lower the FOT, the more orbitals are included in the atomic fragment. In other words, the fragment sizes are systematically increased (Figure 5) to yield increasingly accurate energies (Table 4). We also note that the sizes of the unoccupied spaces are very similar for the MP2 and CCSD calculations, whereas a larger occupied buffer is required for the CCSD case due to a more extensive coupling of amplitudes, in accordance with the locality analysis in Section 4.2.

7.1.5. The δ Parameter. Let us finally comment on the value of the δ parameter, which has been set to 0.01 in all calculations presented above. In Figure 6 we have plotted the error in the total MP2 correlation energy compared to a full molecular MP2 calculation for various FOTs and δ parameters. The error in the total energy decreases by roughly an order of magnitude when decreasing the FOT by an order of magnitude, as discussed in Section 7.1.2, whereas the errors are practically independent of the choice of δ parameter in eq 68 (the lines in Figure 6 are almost horizontal). In general we may therefore control the error in the total correlation energy by tightening the FOT for a fixed value of δ. Even though $\delta = 0.01$ was used in the calculations presented above, even $\delta = 0.1$ seems to be sufficient for a proper description of the orbitals. To understand why a value of $\delta = 0.1$ is sufficient, recall that the atomic fragment orbital space $\{\mathcal{A}_P\}$ is the union of atomic extents for all atoms in the AOS $\mathcal{A}_P$ (see the discussion in Section 6.1). Since the EOS orbitals used for evaluating the atomic fragment energy are assigned to atoms in the central part of the fragment (see Figure 2) and are expanded on all atoms in the atomic fragment extent (including buffer regions), a very accurate description of the EOS orbitals is obtained, even for $\delta = 0.1$. The MOs in the buffer space are described less accurately for $\delta = 0.1$, but this has a very small indirect effect on the atomic fragment energy.

7.2. DEC-MP2 Calculations on Polyalanine Strings. In the previous section we demonstrated that for a given molecule the correlation energy error in a DEC calculation may be controlled by varying the FOT. In this section we investigate how the correlation energy error and the fragment sizes are affected when we systematically increase the molecular size using a fixed FOT.

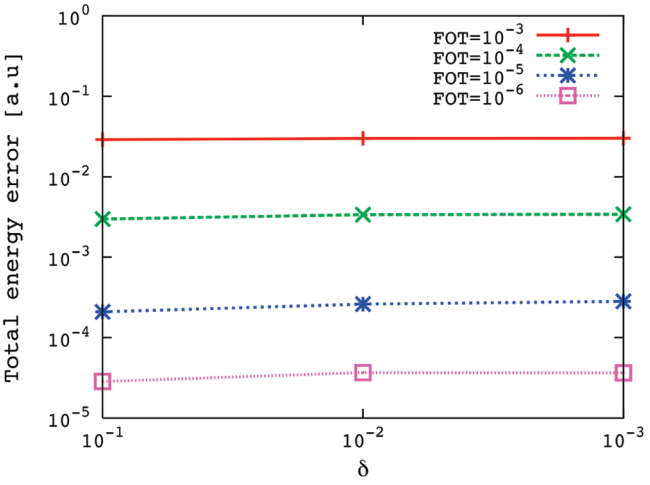

Figure 6. Total energy errors for various fragment optimization thresholds (FOTs) and δ values. In general, the energy error is proportional to the FOT, whereas it is very insensitive to the value of δ. The calculations have been carried out on the $C_{14}H_2$ molecule at the MP2/cc-pVDZ level of theory.

The molecules under investigation are extended peptide chains containing 2, 4, 6, 8, or 10 alanine residues, see Figure 7. The geometries were obtained using extended peptide chains in the Maestro program[41] without carrying out any additional optimizations. We thus have a set of molecular structures where the size is systematically increased. All DEC calculations presented in this section were carried out at the MP2/cc-pVDZ level of theory using a FOT of 10^{-4}, $\delta = 0.1$, orbital step size parameters of 2.0 au, and normalization of the orbitals. The validity of using $\delta = 0.1$ was discussed in Section 7.1.5. We shall not go into any detail about the choice of step size parameter, as we are currently developing a more sophisticated fragment optimization algorithm which does not involve any orbital step size radii.

7.2.1. Energy Errors. In Table 5 we present the correlation energy errors obtained in DEC-MP2 calculations compared to

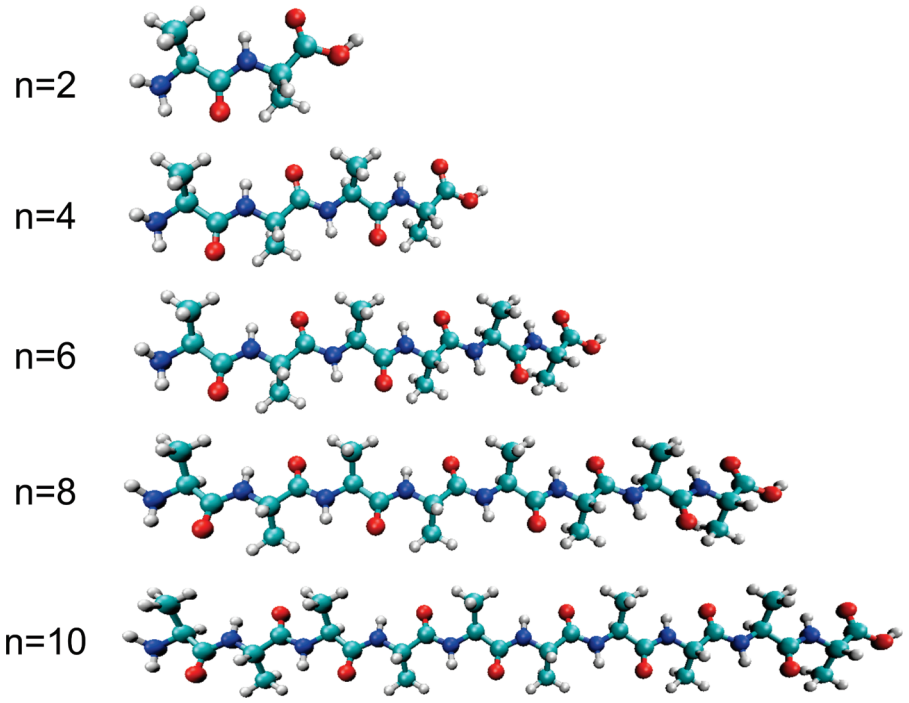

Figure 7. Extended polyalanine structures, and *n* is the number of alanine residues.

full molecular MP2 calculations. As expected the total correlation energy (E_{corr}) increases linearly with molecular size. The absolute energy errors (third column of Table 5) also roughly increase linearly with system size, although slight deviation from this behavior is observed, in particular for the alanine string with eight residues. The fact that the absolute error increases with molecular size can be understood by recalling that the total correlation energy is calculated as a sum of atomic fragment and atomic pair fragment interaction energies, see eq 5. For MP2 calculations we have observed that the fragment energy errors are all positive (see also the discussion in Section 7.1.2), and the errors therefore accumulate instead of canceling each other, such that the total absolute energy error increases with system size. In other words, the absolute MP2 energy errors are positive and size extensive. However, the percentage of E_{corr} that is recovered (fourth column of Table 5) remains roughly constant at approximately 99.93% as we increase the molecular size. The relative energy error is thus independent of the size of the molecule under investigation, i.e., regardless of the system size, the DEC scheme ensures control of the relative energy error. Cluster amplitudes, and therefore also molecular properties (e.g., molecular gradients), are thus determined with about equal accuracy independently of the system size.

7.2.2. Locality of Orbitals and Orbital Spaces. We now consider the sizes of the spaces employed in the DEC-MP2 calculations on the molecules in Figure 7. Let us first consider the locality of the LCM basis[30] which ultimately defines the locality of the DEC calculations. In the fifth column in Table 5 we have plotted the maximum orbital spread for the LCM orbitals employed in the calculation. The orbital spread is a measure of the locality of the orbitals (see ref 30 for a more thorough discussion of orbitals spreads). It is clear that the maximum orbital spread is independent of the molecular size. The locality of the LCM orbitals is manifested in the sizes of the DEC orbital fragments as may be seen from Figure 8, where we

Table 5. Absolute (ΔE_{corr}) and Relative (% E_{corr}) Energy Errors of DEC-MP2/cc-pVDZ Calculations on Polyalanine Strings of Increasing Lengths Compared to Full Molecular Calculations[a]

no. alanines	E_{corr} (au)	ΔE_{corr} (au)	% of E_{corr}	orb. spread (Å)
2	−1.737622	0.0015	99.92	1.91
4	−3.278702	0.0020	99.94	1.92
6	−4.819890	0.0034	99.93	1.90
8	−6.361143	0.0035	99.95	1.89
10	−7.902389	0.0056	99.93	1.95

[a] Also shown is the maximum orbital spread (orb. spread) which is a measure of the locality of the least-change molecular (LCM) basis.

have plotted the average (A) and maximum (B) sizes of the orbital space radii in the DEC atomic fragment calculations as a function of the number of alanine residues. As we increase the size of the molecular system, the average and maximum orbital space radii become independent of the system size, in accordance with the locality analysis in Section 3.2. This is crucial for the applicability of the DEC scheme to calculate correlation energies for large molecular systems: regardless of the total molecular size the individual (independent) fragment calculations are small enough that they may be carried out using standard quantum chemical techniques (see Section 6.3). Thus, provided that a sufficient number of processors is available, the total wall time for carrying out an MP2 calculation for a large molecule system is defined by the largest fragment involved in the calculation.

7.2.3. Sizes of Atomic Fragment Extents. In a DEC fragment calculation the molecular integrals are calculated only for atoms in the atomic fragment extent (see Section 6.1 and Figure 2). In Figure 9 we have plotted the average (green dashed line) and the maximum (blue dotted line) number of atoms in the atomic

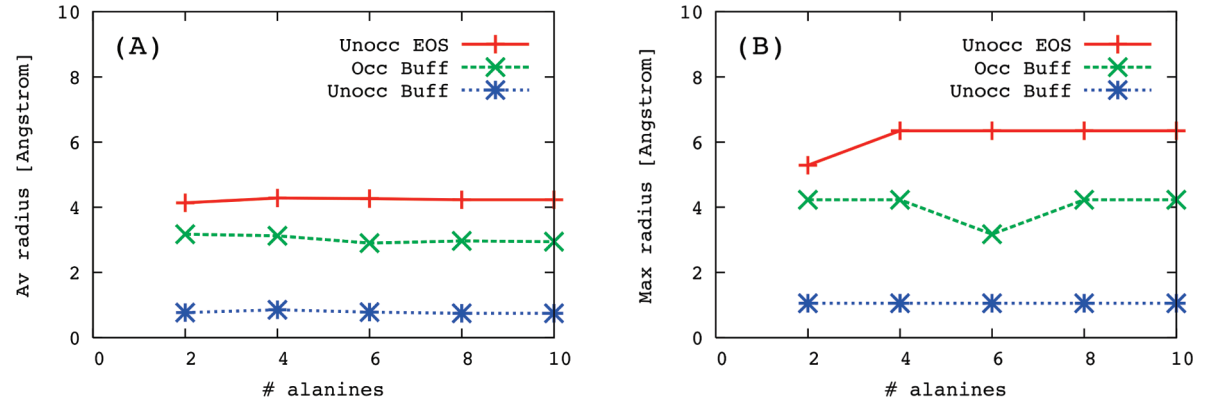

Figure 8. Average (A) and maximum (B) unoccupied energy (red solid curve), occupied buffer (green dashed curve), and unoccupied buffer (blue dotted curve) orbital space radii in DEC atomic fragment calculations as a function of the number of alanine residues.

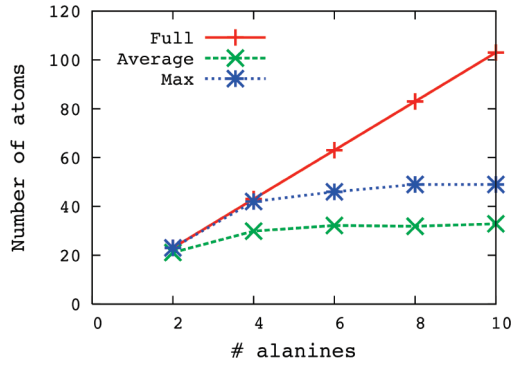

Figure 9. Average (green dashed) and maximum (blue dotted) number of atoms in the atomic fragment extent $\{\mathcal{A}_P\}$ in DEC atomic fragment calculations as a function of the number of alanine residues. For comparison the number of atoms in the corresponding full molecular calculations (red solid) is also plotted.

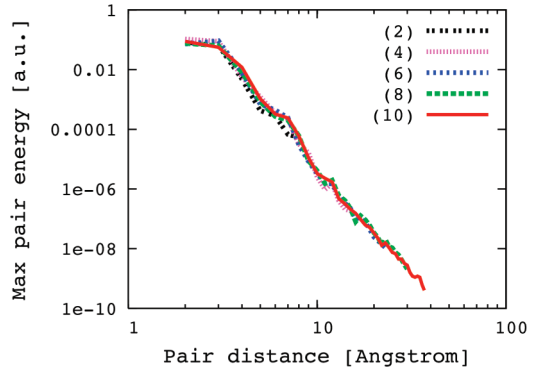

Figure 10. Maximum pair interaction energies (absolute values) against pair distance for polyalanine strings with 2, 4, 6, 8, and 10 alanine residues as indicated in the figure.

fragment extent $\{\mathcal{A}_P\}$ in DEC atomic fragment calculations as a function of the number of alanine residues. For comparison, we have also plotted the total number of atoms in the polyalanine molecules (red solid line). For the smallest polyalanines all atoms in the molecule are included in some of the atomic fragment extents. However, as the molecular size increases, the average and maximum sizes of the atomic fragment extents approach a constant value of roughly 35 and 50 atoms, respectively. This is in line with that we observed in Figure 8: for "large" molecules the fragment sizes are independent of the molecular size.

Table 6. Total Pair Interaction Energy ($\Sigma_{P>Q}\Delta E_{PQ}$) for Different Pair Distance Cutoffs[a]

distance thr. (Å)	$\Sigma_{P>Q}\Delta E_{PQ}$ (au)
all pairs	−2.53084
30.0	−2.53084
25.0	−2.53084
20.0	−2.53084
15.0	−2.53084
10.0	−2.53077
5.0	−2.52383

[a] The calculations were carried out on a polyalanine string with 10 alanine residues at the DEC-MP2/cc-pVDZ level of theory.

7.2.4. Pair Interaction Energies for Polyalanines. Let us also comment on the pair distance dependence of the pair interaction energies ΔE_{PQ} for the polyalanine strings. In Figure 10 we have plotted the maximum pair interaction energies (absolute values) against the pair distance for polyalanine strings with 2, 4, 6, 8, and 10 alanine residues. The fact that the different polyalanine plots almost coincide show that the pair interaction energies, which represent dispersion interactions for larger pair distances, are of similar sizes in all molecules. This is hardly surprising as the molecules are of similar structure, see Figure 7. The magnitude of the pair interaction energies rapidly decreases with pair distance, since dispersion forces ideally depend on the inverse pair distance R_{PQ} to the sixth power.[42] In fact, the (almost) straight line in the double logarithmic plot in Figure 10B does indicate a power dependence, i.e., $\Delta E_{PQ} \propto R_{PQ}^{-n}$. A regression plot for the polyalanine string containing 10 alanine residues yields $n \approx 7.1$. The idealized case where $n = 6$ occurs for two relatively distant atoms and assumes that higher order interactions (i.e., terms proportional to R_{PQ}^{-n} with $n > 6$) can be neglected. What we observe in Figure 10 is a mixture of pair interactions referencing localized orbitals centered on different atom types in different environments and also includes higher order interactions. It is therefore not surprising that our power exponent differs slightly from six.

From a practical point of view the rapid decrease of pair interaction energies with pair distance implies that pairs separated by more than some pair distance cutoff may be omitted from the DEC calculation without affecting the total correlation

energy. To quantify the contributions of distant atomic pairs to the correlation energy, we have in Table 6 given the total pair interaction energy ($\Sigma_{P>Q}\Delta E_{PQ}$) for different values of the pair distance cutoff for a polyalanine string containing 10 alanine residues. It is seen that atomic pairs separated by more than, say, 10 Å do not contribute significantly to the total pair energy considering that the atomic fragment energies were only calculated to a threshold of 10^{-4} au. Pairs separated by more than 10 Å may therefore be omitted from the DEC calculation. Thus, for a large molecule the number of pair interaction energies that needs to be calculated does indeed scale linearly with the size of the molecule.

In a practical calculation, where the total pair energy with all pairs included is not known, a pair distance cutoff of 10 Å seems reasonable for a general nonmetallic molecular system, whereas 15 Å is a more conservative value.

The short version of this discussion is that pair interaction energies describe dispersion effects, which is the same physical effect for all molecules. The threshold for neglecting pair interactions can thus be set based on the distance dependence of this interaction.

8. SUMMARY

We have presented a locality analysis of the MP2 and CCSD amplitude equations which demonstrate that—assuming that a set of local HF orbitals is available, such as the LCM basis,[30] —a full CC calculation can be divided into a set of small independent atomic fragment and atomic pair fragment calculations, where the fragment orbital spaces constitute small subsets of the total orbital space. This is possible because, using a local HF basis, the contributions to the MP2 and CCSD amplitude equations rapidly decrease with increasing spatial distance between the central atom(s) in the atomic (pair) fragment and the orbitals referencing the summation indices in the residual expression.

In a practical calculation, the optimal sizes of the atomic fragment orbital spaces are not known a priori. Rather, they have to be determined during the calculation to ensure that the atomic fragment energies are determined to a preset threshold. Subsequently the atomic pair fragments are formed from unions of atomic fragments to determine the pair interaction energies. A CC calculation on a large molecular system may thus be carried out in terms of CC calculations on small orbital fragments of the total molecular system, where the sizes of the orbital fragments are determined in a black box manner during the calculation to ensure that the total CC correlation energy is calculated to a preset energy threshold.

In summary, a DEC calculation consists of a series of independent atomic fragment calculations followed by a series of independent atomic pair fragment calculations. Pairs which are widely separated in space may be neglected, as their corresponding pair interaction energies (representing dispersion interactions) are exceedingly small. Thus, for large molecules the number of pair fragments with nonvanishing contributions to the correlation energy scales linearly with the number of atoms in the system. The DEC scheme is therefore linearly scaling and embarrassingly parallelizable with roughly the same wall time for small and large molecular systems provided that a sufficient number of processors are available.

Numerical results demonstrate that the energy errors in a DEC-MP2 or DEC-CCSD calculation compared to a full molecular calculation may be defined by the FOT parameter prior to the calculation. We have also demonstrated that the absolute energy errors in a DEC-MP2 calculation compared to a full calculation are size extensive, whereas the relative energy errors are independent of the system size. Thus, in a relative sense, small and large molecules are treated with equal precision in DEC-MP2 calculations. Furthermore, it has been shown that the fragment sizes are independent of the molecular system size and that the fragments are small enough to be treated using standard quantum chemical implementations. Therefore, a DEC correlation energy calculation for a general molecule can be carried out provided that the HF calculation can be carried out and that a local HF basis can be determined.

AUTHOR INFORMATION

Corresponding Author

*E-mail: kasperk@chem.au.dk.

Present Addresses

[‡]Department of Chemistry, Northwestern University, Evanston, Illinois 60208−3113, United States

[§]Department of Chemistry, Oslo University, P.O. Box 1033, Blindern, 0315 Oslo, Norway

ACKNOWLEDGMENT

This work has been supported by the Lundbeck Foundation and the Danish Center for Scientific Computing (DCSC).

REFERENCES

(1) Shavitt, I.; Bartlett, R. J. *Many-Body Methods in Chemistry and Physics. MBPT and Coupled-Cluster Theory*; Cambridge University Press: Cambridge, U.K., 2009.
(2) Hartree, D. R. *Proc. Cambridge Philos. Soc.* **1928**, *24*, 89.
(3) Fock, V. *Z. Phys.* **1930**, *61*, 126.
(4) Pulay, P. *Chem. Phys. Lett.* **1983**, *100*, 151.
(5) Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 296.
(6) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457–464.
(7) Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916.
(8) Subotnik, J. E.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2005**, *123*, 114108.
(9) Saebø, S.; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213–236.
(10) Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1996**, *104*, 6286–6297.
(11) Schütz, M.; Hetzer, G.; Werner, H.-J. *J. Chem. Phys.* **1999**, *111*, 5691–5705.
(12) Schütz, M. *J. Chem. Phys.* **2000**, *113*, 9986.
(13) Schütz, M.; Werner, H.-J. *J. Chem. Phys.* **2001**, *114*, 661–681.
(14) Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330–8343.
(15) Christiansen, O.; Manninen, P.; Jørgensen, P.; Olsen, J. *J. Chem. Phys.* **2006**, *124*, 084103.
(16) Weijo, V.; Manninen, P.; Jørgensen, P.; Christiansen, O.; Olsen, J. *J. Chem. Phys.* **2007**, *127*, 074106.
(17) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *125*, 074116.
(18) Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *121*, 10935–10944.
(19) Li, S.; Ma, J.; Jiang, Y. *J. Comput. Chem.* **2001**, *23*, 237.
(20) Li, W.; Piecuch, P.; Gour, J. R.; Li, S. *J. Chem. Phys.* **2009**, *131*, 114109.
(21) Li, W.; Piecuch, P. *J. Phys. Chem. A* **2010**, *114*, 8644.
(22) Kobayashi, M.; Nakai, H. *J. Chem. Phys.* **2008**, *129*, 044103.
(23) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.
(24) Stoll, H. *Chem. Phys. Lett.* **1992**, *191*, 548.

(25) Friedrich, J.; Hanrath, M.; Dolg, M. *J. Chem. Phys.* **2007**, *126*, 154110.
(26) Almlöf, J. *Chem. Phys. Lett.* **1991**, *181*, 319.
(27) Ayala, P. Y.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 3660–3671.
(28) Lambrecht, D. S.; Doser, B.; Ochsenfeld, C. *J. Chem. Phys.* **2005**, *123*, 184102.
(29) Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S. *Reviews in computational chemistry*; VCH Publishers, Inc.: New York, 2007; Vol. 23, p 1.
(30) Ziółkowski, M.; Jansík, B.; Jørgensen, P.; Olsen, J. *J. Chem. Phys.* **2009**, *131*, 124112.
(31) Ziółkowski, M.; Jansík, B.; Kjærgaard, T.; Jørgensen, P. *J. Chem. Phys.* **2010**, *133*, 014107.
(32) Boughton, J. W.; Pulay, P. *J. Comput. Chem.* **1993**, *14*, 736.
(33) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley & Sons: Chichester, U.K., 2002.
(34) Hetzer, G.; Pulay, P.; Werner, H.-J. *Chem. Phys. Lett.* **1998**, *290*, 143–149.
(35) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085.
(36) Goedecker, S.; Scuseria, G. E. *Comp. Sci. Eng.* **2003**, *5*, 14.
(37) Polly, R.; Werner, H.-J.; Manby, F. R.; Knowles, P. J. *Mol. Phys.* **2004**, *102*, 2311.
(38) Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S. In *Reviews in computational chemistry*; Lipkowitz, K. B., Cundari, T. R., Eds.; VCH Publishers, Inc.: New York, 2007, Vol. 23.
(39) Usvyat, D.; Schütz, M. *Theo. Chem. Acc.* **2005**, *114*, 7.
(40) Ziółkowski, M.; Weijo, V.; Jørgensen, P.; Olsen, J. *J. Chem. Phys.* **2008**, *128*, 204105–12.
(41) *Maestro*, v. 8.5; Schrödinger, LLC: Cambridge, MA, 2008; http://www.schrodinger.com.
(42) Sakurai, J. *J. Modern Quantum Mechanics*, revised ed.; Addison-Wesley Publishing Company: New York, 1994.