3-2015

Combined Fragment Molecular Orbital Cluster in Molecule Approach to Massively Parallel Electron Correlation Calculations for Large Systems

Alexander D. Findlater
Iowa State University, adf@iastate.edu

Federico Zahariev
Iowa State University, fzahari@iastate.edu

Mark S. Gordon
Iowa State University, mgordon@iastate.edu

Combined Fragment Molecular Orbital Cluster in Molecule Approach to Massively Parallel Electron Correlation Calculations for Large Systems

Abstract
The local correlation "cluster-in-molecule" (CIM) method is combined with the fragment molecular orbital (FMO) method, providing a flexible, massively parallel, and near-linear scaling approach to the calculation of electron correlation energies for large molecular systems. Although the computational scaling of the CIM algorithm is already formally linear, previous knowledge of the Hartree–Fock (HF) reference wave function and subsequent localized orbitals is required; therefore, extending the CIM method to arbitrarily large systems requires the aid of low-scaling/linear-scaling approaches to HF and orbital localization. Through fragmentation, the combined FMO-CIM method linearizes the scaling, with respect to system size, of the HF reference and orbital localization calculations, achieving near-linear scaling at both the reference and electron correlation levels. For the 20-residue alanine α helix, the preliminary implementation of the FMO-CIM method captures 99.6% of the MP2 correlation energy, requiring 21% of the MP2 wall time. The new method is also applied to solvated adamantine to illustrate the multilevel capability of the FMO-CIM method.

Disciplines
Chemistry

Comments
Reprinted (adapted) with permission from *Journal of Physical Chemistry A* 119 (2015): 3587, doi:10.1021/jp509266g.

THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

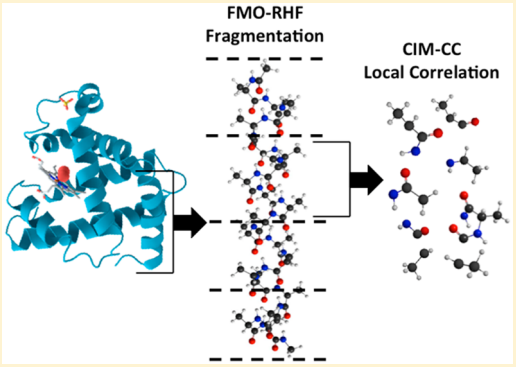
pubs.acs.org/JPCA

Combined Fragment Molecular Orbital Cluster in Molecule Approach to Massively Parallel Electron Correlation Calculations for Large Systems.

Alexander D. Findlater, Federico Zahariev, and Mark S. Gordon*

Department of Chemistry and Ames Laboratory, Iowa State University, Spedding Hall, Ames, Iowa 50014, United States

ABSTRACT: The local correlation "cluster-in-molecule" (CIM) method is combined with the fragment molecular orbital (FMO) method, providing a flexible, massively parallel, and near-linear scaling approach to the calculation of electron correlation energies for large molecular systems. Although the computational scaling of the CIM algorithm is already formally linear, previous knowledge of the Hartree−Fock (HF) reference wave function and subsequent localized orbitals is required; therefore, extending the CIM method to arbitrarily large systems requires the aid of low-scaling/linear-scaling approaches to HF and orbital localization. Through fragmentation, the combined FMO-CIM method linearizes the scaling, with respect to system size, of the HF reference and orbital localization calculations, achieving near-linear scaling at both the reference and electron correlation levels. For the 20-residue alanine α helix, the preliminary implementation of the FMO-CIM method captures 99.6% of the MP2 correlation energy, requiring 21% of the MP2 wall time. The new method is also applied to solvated adamantine to illustrate the multilevel capability of the FMO-CIM method.



I. INTRODUCTION

It is well-documented that the treatment of electron correlation is necessary for the reliable prediction of most molecular properties and observables. Hierarchies of configuration interaction (CI), many-body perturbation theory (MBPT), and coupled-cluster (CC) theory post-Hartree-Fock (HF) methods have been developed to model the effects of electron correlation; however, most of these methods become intractable with increasing system size due to the prohibitable scaling of computation time and resources. (The simplest electron correlation methods scale to the fifth power.)

The unfavorable computational scaling of electron correlation methods is a consequence of the generally delocalized nature of canonical HF molecular orbitals (MOs).[1−7] Electron correlation is a local phenomenon in nonmetallic systems; however, the use of a delocalized MO basis in post-HF methods makes this locality difficult to exploit as (1) delocalized MOs prevent weakly correlated electron pairs (weakly interacting, spatially separated pairs) from being ignored in a systematic way and (2) the dimensionality of the delocalized virtual MO basis required to accurately describe electron correlation increases at an unphysical and computationally prohibitive rate with molecular size.[3−7] The disconnect between the locality of the correlation interaction and the high scaling of post-HF correlation methods has been bridged by many methodologies over the last three decades, including local correlation approaches based on atomic orbital (AO) and localized molecular orbital (LMO) domains and energy-based fragmentation methods. (See refs 8−10 for reviews on both local correlation and fragmentation methods.)

The seminal local correlation formalism of Pulay employs nonorthogonal projected atomic orbitals (PAOs) to construct inherently local virtual orbital domains for each occupied LMO pair while ignoring (or approximating) contributions from weakly interacting (spatially separated) LMO pairs. This formalism was successfully applied by Pulay and Saebø to configuration interaction singles and doubles (CISD),[3] the coupled-electron pair approximation (CEPA),[4] second-order perturbation theory (MP2/MBPT2),[5] and fourth-order perturbation theory (MBPT4).[6,7] On the basis of the self-consistent electron-pair (SCEP) theory of Meyer,[11] the nonorthogonal virtual AOs localized on only a few atoms will contribute to the correlation of two significantly correlated LMOs, reducing the computational scaling to near-linear because the size of the virtual orbital domain for each LMO pair is then independent of the total system size.[3−7]

The formalism of Pulay and Saebø was later generalized by Werner, Schütz, and coworkers and extended to coupled-cluster singles and doubles (CCSD) and the noniterative triples correction (CCSD(T)), achieving low-order scaling,[12] and with integral-direct methods, true linear scaling[13−17] of computation time, and resources with respect to system size. Various alternative localized orbital approaches to local CC include, but

Received: September 12, 2014
Revised: March 19, 2015
Published: March 20, 2015

DOI: 10.1021/jp509266g
J. Phys. Chem. A 2015, 119, 3587−3593

are not limited to, the orbital-specific-virtual local CC method (OSV-CC) of Yang, Chan, Manby, Schütz, and Werner,[18] the natural linear scaling (NLS) CC method of Flocke and Bartlett,[19] the pair natural orbital CC method of Neese et al.,[20] the CC divide-and-conquer (DC-CC) formalism of Kobayashi and Nakai,[21] and the divide-expand-consolidate (DEC) method of Jorgensen et. al.[22]

Of interest in the current work is the single-environment cluster-in-molecule (CIM) local correlation formalism of Li et al.[23−25] and its extension by Li and Piecuch et al.[26−29] to the highly vectorized MP2, CCSD,[30] CCSD(T)[31] and size-extensive left eigenstate completely renormalized CR-CC-(2,3)[32]-coupled cluster codes of Piecuch in GAMESS (General Atomic and Molecular Electronic Structure System);[33] the CR-CC(2,3) and linear-scaling CIM-CR-CC(2,3) methods are capable of accurately describing biradicals and single-bond breaking regions of a potential energy surface.[27,34] After reference HF and subsequent Boys[35] localization calculations have been performed, the CIM algorithm[26−29] constructs orthogonal occupied and virtual correlating orbital domains for each LMO. Occupied LMOs are projected onto the AO basis assigned to each orbital domain and virtual orbitals are orthogonally constructed from a basis of PAOs localized on the atom centers of the truncated set of AO basis functions. (More details of the CIM algorithm are given in the next section, and a complete description is given in refs 24−29.) The size of each correlating orbital domain is independent of the total system size and leads to an implementation with which computation time and resources scale linearly with system size. The occupied and virtual orbital spaces of each orbital domain are separately self-canonicalized resulting in quasi-canonical molecular orbitals (QCMO) to take advantage of the vectorized MP2 and CC canonical orbital codes in GAMESS.[26−29]

As an alternative to intrinsically local LMO and AO orbital-domain methods, energy-based fragmentation methods explicitly subdivide a large system, at the level of atoms, into smaller but chemically sensible fragments. While most of the LMO and AO approaches to local correlation previously mentioned require a previous knowledge of the HF reference MOs and the subsequent LMOs of the entire system, it is typical of fragmentation methods to calculate both the HF reference and correlation energy for each fragment and assemble the total correlation energy from fragment contributions. The details of this procedure vary widely among different fragmentation methods, as reviewed by Gordon, Fedorov, Pruitt, and Slipchenko.[10]

The fragment molecular orbital (FMO) method, proposed by Kitaura et al.,[36] utilizes a many-body expansion to decompose the total energy of a large system, fragmented into 1- and 2-body (FMO2) and optionally 3-body (FMO3)[37] energy contributions. Higher order many-body effects are approximately treated by the inclusion of an electrostatic potential (a Coulomb potential generated by the other fragments) in the Fock operator of each monomer (1-body), dimer, and trimer calculation.[38] In addition to HF, the FMO method is already compatible with the wide range of conventional correlation methods found in GAMESS, including density functional theory (DFT),[39] MP2,[40] CI, and multi-reference CI,[41] CCSD, and CCSD(T).[42]

Monomer, dimer, and trimer energy/property calculations in the FMO method are independent of one another, while the size of each monomer is independent of the total system size. Computations of dimers and trimers formed by spatially separated fragments are approximated with electrostatic potentials. As a result, the FMO method in practice is a massively parallel, nearly linear scaling energy fragmentation approach to both HF and post-HF energies and properties. Of course, fragmentation methods, including FMO, are but one approach to linear scaling HF.[10]

While CIM and other LMO/AO-based local correlation methods scale linearly with system size, the prerequisite full-system HF calculation scales to the fourth power in naive implementations and, with prescreened integral-direct methods, between the second and third power for large molecular species. Consequently, the extension of linear-scaling local correlation methods to arbitrarily large systems will require the aid of low-scaling/linear-scaling implementations of the HF and localization steps. With this motivation in mind, the current work presents the combined fragment molecular orbital and cluster-in-molecule (FMO-CIM) method, a massively parallel and near linear scaling approach for the treatment of the electron correlation of arbitrarily large systems.

FMO increases the size of systems that can be treated using the CIM method by linearizing the scaling of the prerequisite HF and localization calculations; CIM also allows users of the FMO method more potential flexibility when choosing an FMO fragmentation scheme. The large difference in scaling, with respect to FMO fragment size, between HF and high-level correlation methods sometimes makes it difficult to simultaneously design a chemically and computationally optimal fragmentation scheme. By replacing conventional MP2 and CC correlation methods in FMO with their linear scaling CIM-MP2 and CIM-CC variants, fragmentation schemes can be optimized for the lower scaling HF and localization calculations.

In the following, the FMO-CIM methodology and the implementation in GAMESS are described. Correlation energy and wall-time comparisons are made between CIM/MP2, FMO-CIM/MP2, and MP2 (in GAMESS) for the 20-residue alanine α helix test case. In addition, the multilayer FMO-CIM/CR-CC(2,3) method is applied to solvated adamantane to demonstrate the flexibility of the FMO-CIM approach to treat different portions of a system with varying levels of theory.

II. OVERVIEW OF THE FRAGMENT MOLECULAR ORBITAL AND CLUSTER-IN-MOLECULE METHOD

A. Methodology. The FMO-RHF (restricted Hartree−Fock) energy of a molecular system divided into N fragments is defined by the following many-body expansion (eq 1)

$$E^{\mathrm{FMO\text{-}RHF}} = \sum_{i}^{N} E_i^{\mathrm{RHF}} + \sum_{i>j}^{N} (E_{ij}^{\mathrm{RHF}} - E_i^{\mathrm{RHF}} - E_j^{\mathrm{RHF}}) + \sum_{i>j>k}^{N} \begin{bmatrix} (E_{ijk}^{\mathrm{RHF}} - E_i^{\mathrm{RHF}} - E_j^{\mathrm{RHF}} - E_k^{\mathrm{RHF}}) - \\ (E_{ij}^{\mathrm{RHF}} - E_i^{\mathrm{RHF}} - E_j^{\mathrm{RHF}}) - \\ (E_{ik}^{\mathrm{RHF}} - E_i^{\mathrm{RHF}} - E_k^{\mathrm{RHF}}) - \\ (E_{jk}^{\mathrm{RHF}} - E_j^{\mathrm{RHF}} - E_k^{\mathrm{RHF}}) \end{bmatrix} + \cdots \quad (1)$$

In eq 1, E_i, E_{ij}, and E_{ijk} are one-, two-, and three-body energies. n-body refers to the number of fragments (n) explicitly included in the RHF energy calculation for a specific term. The FMO2-RHF and FMO3-RHF methods include up to dimer (two-body) and trimer (three-body) RHF calculations, respectively. An electrostatic embedding potential (ESP)

DOI: 10.1021/jp509266g
J. Phys. Chem. A 2015, 119, 3587−3593

representative of the electron-density and nuclei of the remaining fragments is included in the Fock operator of each n-body RHF calculation to capture higher order many-body effects (eq 2).[38]

$$V_{\mu\upsilon}^{X} = \sum_{K\neq X}^{N}\left\{\sum_{A\in K}\langle\mu|\frac{Z_A}{|r-R_A|}|\upsilon\rangle + \sum_{\rho\sigma\in K} D_{\rho\sigma}^{K}(\mu\upsilon|\rho\sigma)\right\} \tag{2}$$

Lowercase Greek letters run over AOs, Z_A and R_A are, respectively, the charge and position of nucleus, A, and $D_{\rho\sigma}^{X}$ is the electron density matrix for fragment X, where $X = i$ for monomers and $X = ij$ for dimers. The ESP is a function of the electron density of every fragment; therefore, FMO-RHF monomer energies are iterated to self-consistency before dimer/trimer or electron correlation calculations are performed. Calculations involving fragments separated by more than R_{cut} can be approximated by electrostatics to grow the number of dimer and trimer calculations linearly with system size; this is called the separated dimer/trimer approximation. Default values of R_{cut} can be over-ridden by a user.

The FMO total energy (eq 3) is defined as the sum of the FMO-RHF reference and correlation energies.

$$E^{\text{total}} = E^{\text{FMO-RHF}} + E^{\text{FMO-corr}} \tag{3}$$

In the FMO-CIM method, the many-body expansion of the FMO correlation energy is expressed in terms of n-body correlation energies, each calculated with one of the linear scaling CIM-MP2 or CIM-CC (eq 4) methods (i.e., CIM-CCSD, CCSD(T), or CR-CC(2,3)).

$$E^{\text{FMO-corr}} = \sum_{i}^{N} E_i^{\text{CIM-CC}} + \sum_{i>j}^{N}(E_{ij}^{\text{CIM-CC}} - E_i^{\text{CIM-CC}} - E_j^{\text{CIM-CC}}) + \cdots \tag{4}$$

The CIM-CC (or CIM-MP2) correlation energy (eq 5) for fragment X ($X = i$ for monomers, $X = ij$ for dimers) is written in terms of contributions δE_i^{CC} from each occupied valence LMO localized on X, obtained via Boys localization[35] after the FMO-RHF calculation on X.

$$E_X^{\text{CIM-CC}} = \sum_{i}^{\text{LMO}\in X} \delta E_i^{\text{CC}} \tag{5}$$

For CCSD (eq 6), δE_i^{CCSD} is expressed in terms of one- and two-electron integrals (f_i^a and v_{ij}^{ab}, respectively), together with one- and two-body cluster amplitudes (t_a^i and τ_{ab}^{ij}), with ij and ab representing occupied and virtual orbitals, respectively. For the CR-CC(2,3) triples correction (eq 7), $\delta E_i^{2,3}$ is expressed in terms of de-excitation amplitudes and triply excited moments (I_{ijk}^{abc} and M_{abc}^{ijk}).

$$\delta E_i^{\text{CCSD}}(\{P_i^X\}) = \sum_{a\in\{P_i^X\}} f_i^a t_a^i + \frac{1}{2}\sum_{j,a>b\in\{P_i^X\}} v_{ij}^{ab}\tau_{ab}^{ij} \tag{6}$$

$$\delta E_i^{2,3}(\{P_i^X\}) = \frac{1}{3}\sum_{j>k,a>b\in\{P_i^X\}} I_{ijk}^{abc}\text{M}_{abc}^{ijk} \tag{7}$$

The occupied and virtual orbital indices on the right-hand side of eqs 6 and 7 are restricted to those orbitals that are included in a given CIM subsystem $\{P_i^X\}$, where $\{P_i^X\}$ is the set of occupied and virtual correlating orbitals for LMO ϕ_i^X.

The occupied LMOs ϕ_j^X that are included in $\{P_i^X\}$ are those that satisfy eq 8 for ϕ_i^X. In other words, the Fock matrix integral connecting ϕ_j^X with ϕ_i^X must be greater than the cutoff parameter ζ whose default value is 0.003 hartree.

$$\langle\phi_i^X|\hat{F}|\phi_j^X\rangle > \xi \tag{8}$$

Each occupied LMO ϕ_i^X on fragment X is assigned an AO domain $\Omega^X(i)$ that includes the basis functions centered on atoms that have a "significant" Mulliken population contributing to ϕ_i^X. Atoms are added, in order of decreasing populations, until the total orbital population reaches 1.98. The definition of "significant" is represented by a default parameter that can be over-ridden by user input. The AO domain of $\{P_i^X\}$, $\Omega^X(P_i^X)$, is defined as the union of the AO domains assigned to each LMO in $\{P_i^X\}$. The LMOs that satisfy eq 8 are then projected onto $\Omega^X(P_i^X)$.

Equation 9 defines the localized orthonormal virtual orbitals used in $\{P_i^X\}$. $C_{\rho a}$ is determined by canonicalization of a projected atomic-orbital (PAO) basis (eq 10) constructed using $\Omega^X(P_i^X)$. In eq 10, $|\chi_\mu{}^X\rangle \in \Omega^X(P_i^X)$ and i' runs over all $n_o\{P_i^X\}$ occupied LMOs in the subsystem.

$$\phi_a^X = \sum_{\rho\in\Omega^X(P_i^X)} C_{\rho a}\overline{\chi}_\rho^X \tag{9}$$

$$|\overline{\chi}_\rho^X\rangle = (1 - \sum_{i'=1}^{n_o\{P_i^X\}} |\bar{\phi}_\rho^X\rangle\langle\bar{\phi}_{i'}^X|)|\chi_\mu^X\rangle \tag{10}$$

Once the occupied and virtual orbital subsystems $\{P_i^X\}$ have been constructed for each occupied LMO ϕ_i^X on fragment X and the totally overlapping subsystems on X have been combined and cleared of redundancies, an independent MP2 or CC calculation can be performed on each subsystem. Then, the fragment correlation energy is evaluated using eq 5. In practice, LMO domains are canonicalized to form quasi-canonical molecular orbitals (QCMO). For a separate, detailed description of the FMO and CIM algorithms in GAMESS, see refs 36−42 and 26−28, respectively.

B. FMO−CIM Parallel Implementation Using the General Distributed Data Interface. In GAMESS, the FMO-RHF method utilizes the general distributed data interface (GDDI)[43] to achieve a two-level hierarchical parallelization scheme. The upper level (coarse-grained) parallelism is achieved by assigning each independent n-body calculation to a GDDI group that consists of a user-defined number of compute nodes. The lower level (fine-grained) parallelism is then the result of task parallelization (threading) within each GDDI group. The combined use of GDDI and the separated dimer/trimer approximation produces a near-linear-scaling FMO-RHF implementation.

The FMO-CIM method follows a similar parallelization scheme. Following each RHF n-body calculation (in the field of the self-consistently converged ESP), the GDDI group structure is retained for use by the CIM algorithm. Boys LMOs and subsequent CIM subsystems are generated for each n-body fragment (up to dimers for FMO2-CIM and trimers for FMO3-CIM), followed by the parallel execution of each subsystem MP2 or CC calculation within each fragment-assigned GDDI group. Consequently, the composite FMO-CIM method in GAMESS is massively parallel and achieves near-linear scaling with respect to system size, at both the RHF and CIM-CC or CIM-MP2 levels of theory. The FMO-CIM

DOI: 10.1021/jp509266g
J. Phys. Chem. A 2015, 119, 3587−3593

Table 1. Details of the FMO2-CIM/MP2 and CIM/MP2 Calculations[a]

method	ζ (hartree)	number of correlated monomers	number of correlated dimers	average number of MP2 subsystems per monomer	average number of MP2 subsystems per dimer	total number of MP2 subsystems
FMO2-CIM/MP2	0.003	10	17	8	14	318
		5	4	16	36	224
	0.001	10	17	2	8	156
		5	4	20	39	256
CIM/MP2	0.003					97
	0.001					97

[a]The number of correlated monomers was chosen to equal the total number of monomer fragments. Dimers separated by a distance greater than R_{cut} (2 Å) are approximated by electrostatics, while the remaining dimers are treated with CIM-MP2.

method allows the CIM method to be effectively extended to large systems due to the fragmentation and linearization of the reference calculation.

While the FMO-CIM correlation calculations using GDDI reach near-linear scaling with respect to system size, the CIM subsystem generation algorithm, used for calculations in this paper, is *not* parallel within a GDDI group (no fine-grain parallelism). While the wall time and computational resources for the subsystem generation calculation scale linearly with system size, the implementation suffers from a larger prefactor due to the large number of subsystems that must be generated with serial computations within each GDDI group. Parallelization of the subsystem generation code will be addressed in future work.

III. PRELIMINARY CALCULATIONS AND DISCUSSION

A. Alanine Alpha Helix Test Case. To determine the accuracy and performance of the combined FMO-CIM approach relative to standard CIM and canonical correlation methods, FMO2-CIM/MP2, CIM/MP2, and full MP2 (using GAMESS), were applied to the calculation of the electron correlation energy of a 20-residue polyalanine α helix, Ala_{20} (using the FMO2-RHF equilibrium geometry). To study the dependence of the FMO2-CIM/MP2 correlation energy and computational wall time with respect to both the FMO fragment size and the CIM subsystem parameter ζ, calculations were performed using two and four alanine residues per FMO fragment (five- and ten-fragment schemes) while also varying the subsystem size parameter ζ from 0.003 to 0.001 hartree. Recall that a smaller value of ζ corresponds in general to fewer (and therefore larger) subsystems. As ζ approaches zero, only one CIM subsystem that spans the entire FMO fragment will be generated. Then, all LMO pairs on the fragment are correlated together using the entire virtual orbital space. $\zeta = 0$ is equivalent to the FMO-MP2 method.

FMO2-CIM/MP2, CIM/MP2, and MP2 correlation energies were calculated using the 6-31G(d) basis set (1789 total basis functions) on 256 CPU cores (2.0 Ghz Intel E5 2650). For FMO2-CIM/MP2, 16 GDDI groups were defined: Each fragment calculation is assigned to one group, each containing 16 Intel E5 2650 cores. The MP2 subsystem calculations, using the entirety of the 256 cores (one GDDI group), were performed after the FMO2-CIM calculation had generated subsystems for every monomer and correlated dimer; dimers were separated by less than R_{cut}, set to 2 Å. The MP2 subsystem calculations were performed in parallel using one core for each subsystem. The FMO2-CIM/MP2 and CIM/MP2 methods both generated 256 or fewer total subsystems for Ala_{20}, except in the case of the ten-fragment, ζ = 0.001 FMO2-CIM/MP2 calculation. (See Table 1.). The subsystem generation algorithm used in the current work closely resembles that which is in the current version of GAMESS. The recent algorithmic developments of Li et. al present an interesting possibility to further improve the efficiency of the CIM algorithm.[44]

Table 2 contains the RHF reference and correlation energies (in hartree) for MP2, CIM/MP2, and the four FMO2-CIM/MP2 combinations of fragment size and subsystem parameter ζ for Ala_{20}. The percentage of correlation energy recovered and the absolute error (hartree) of FMO2-CIM/MP2 and CIM-MP2 relative to MP2 is also provided.

Table 2. FMO2-CIM/MP2, CIM/MP2, and MP2 Reference and Correlation Energies (Hartree) for Ala_{20}[a]

method	ζ (hartree)	correlation energy (hartree)	correlation energy recovered relative to full MP2	absolute-error relative to full MP2 (hartree)
10 fragments FMO2-CIM/MP2	0.003	−15.219	99.11%	0.14
	0.001	−15.288	99.56%	0.07
5 fragments FMO2-CIM/MP2	0.003	−15.218	99.11%	0.14
	0.001	−15.286	99.55%	0.07
full RHF+CIM/MP2	0.003	−15.216	99.09%	0.14
	0.001	−15.286	99.55%	0.07
MP2	$\zeta \rightarrow 0$	−15.355		

[a]FMO2-CIM/MP2 and CIM/MP2 percentage of correlation energy recovered and absolute error (hartree) are relative to full MP2 on the entire system.

With ζ set to 0.001 hartree, CIM/MP2 and FMO2-CIM/MP2 using the ten- and five-fragment schemes both capture ∼99.6% of the total MP2 correlation energy with absolute errors of ∼0.07 hartree. With ζ constant, changing the FMO2-CIM/MP2 fragment size from two to four residues has little effect on the total correlation energy that is recovered; however, decreasing the subsystem parameter ζ from 0.003 to 0.001 hartree leads to a 2-fold reduction in the FMO2-CIM/MP2 and CIM/MP2 absolute error relative to MP2. The FMO2-CIM/MP2 method reproduces the CIM/MP2 method for both fragmentation schemes. Both the FMO-CIM/MP2 and CIM-MP2 methods, with ζ set to 0.001 hartree, reasonably reproduce the full MP2 result, as ∼99.6% of the correlation energy is captured. The FMO2-CIM/MP2 correlation energy is much more sensitive to the value of the subsystem size parameter ζ than to the size of the underlying FMO fragments, and further gains in correlation energy can be achieved by further decreasing the value of the subsystem parameter ζ (increasing subsystem size). The 10-fragment and 5-fragment

DOI: 10.1021/jp509266g
J. Phys. Chem. A 2015, 119, 3587−3593

Table 3. FMO2-CIM/MP2 Calculation Wall Times (min) Compared with CIM/MP2 Calculations on the Entire Ala_{20} System[a]

method	reference calculation wall time (min)	ζ (hartree)	Boys localization and CIM subsystem generation (min)	average MP2 calculation time (min)	largest MP2 calculation time (min)	total wall time (min)
10 fragments FMO2-CIM/MP2	2.5	0.003	16.7	0.9	3.4	25.6
		0.001	19.2	2.0	9.3	31.0
5 fragments FMO2-CIM/MP2	12.4	0.003	174.8	1.8	5.8	193.0
		0.001	222.4	5.7	25.4	260.2
full RHF+CIM/MP2	24.7	0.003	843.7	2.1	7.2	875.6
		0.001	773.2	7.6	17.6	815.5
full RHF+MP2	24.7				126.4	151.1

[a]Times spent calculating the RHF and FMO-RHF reference and generating the CIM subsystems are given separately as well as the total wall time. The 10- and 5-fragment FMO2-CIM/MP2 schemes include two and four alanine residues per fragment, respectively. Calculations were performed on 256 2.0 Ghz Intel E5 2650 cores across 16 GDDI groups.

Table 4. MFMO3-CIM/CR-CC(2,3) (ζ = 0.003) and MFMO3/CR-CC(2,3) Correlation Energies (Hartree) and Total Wall Times (min) for Water-Solvated Adamantane[a]

method	MP2 correlation energy (Hartree)	CCSD correlation energy (Hartree)	CR-CC(2,3) triple correction (Hartree)	total wall time (min)
MFMO3-CIM/CR-CC(2,3)	−1.305	−1.4029	−1.4528	187.5
	(−99.19%)	(−99.56%)	(99.51%)	
MFMO3/CR-CC(2,3)	−1.3156	−1.409	−1.460	536.3

[a]Percentage of MFMO/CR-CC(2,3) correlation energy captured by MFMO-CIM/CR-CC(2,3) is given in parentheses.

FMO2 schemes reproduce the full RHF reference energy to 3 and 0.2 millihartrees respectively, which is consistent with accuracies achieved in previous FMO2 studies.[39,40]

Only 21% of the full RHF+MP2 total wall time (Table 3) is required to capture 99.56% of the MP2 correlation energy using the ten-fragment FMO2-CIM/MP2 (ζ = 0.001 hartree) calculation, a difference of 120 min. The full RHF+MP2 wall time is the combined wall time for the RHF and MP2 calculations. Using the five-fragment scheme (four alanine residues per fragment), the reference FMO2-RHF wall time increases by a factor of 5, while the CIM subsystem generation step time increases by over an order of magnitude. The five-fragment scheme yields very little improvement in correlation energy relative to the ten-fragment scheme, despite the 8-fold increase in total wall time.

The five-fragment FMO2-CIM/MP2 calculations and both of the CIM/MP2 calculations have longer wall times compared with full MP2 due to the serial implementation of the subsystem generation step. The Boys localization step-time is a small percentage of the total wall time. The CIM subsystem generation step and subsequent MP2 subsystem calculations scale linearly with total system size; however, there is a large prefactor bottleneck associated with generating a large number of subsystems in serial for each FMO fragment. Although the ten-fragment FMO-CIM/MP2 calculation produces the most subsystems, the size of the fragments is small enough to limit the number of CIM subsystems per GDDI group, allowing for a shorter wall time compared with full MP2. Subsystem generation is more efficiently parallelized with GDDI when using smaller fragments. Future work will address the parallelization of the subsystem generation step in the CIM and FMO-CIM methods and will elucidate the nontrivial RHF bottleneck of the CIM procedure.

B. Multilayer FMO-CIM Applied to Water-Solvated Adamantane. To demonstrate the multilayer FMO-CIM method, the MFMO3-CIM/CR-CC(2,3) electron correlation energy was calculated and compared with MFMO3-CIM/CR-CC(2,3) for the adamantane nanodiamond ($C_{10}H_{16}$) solvated by 150 waters. The solvated adamantane structure is a snapshot from an equilibrated combined quantum-mechanics-effective fragment potential (EFP[45]), QM/EFP molecular dynamics simulation trajectory using the 6-31G(d) basis set, and the RHF level of QM theory. The multilayer functionality of MFMO-CIM facilitates the calculation of CIM correlation energies on targeted fragments. In this case, the MFMO3-CIM/CR-CC(2,3) correlation energy is calculated only for the nanodiamond (solute), while the 150 waters (solvent) are treated at the FMO3-RHF level. The three-body corrections in FMO3-CIM are necessary to account for the three-body nature of the hydrogen-bonding interactions in water. MFMO3-CIM/CR-CC(2,3) correlation calculations were performed using the 6-31G(d) basis set (3032 total basis functions) on 256 2.0 Ghz Intel E5 2650 cores spread across 16 GDDI groups.

Table 4 contains the correlation energy and wall-time comparisons between MFMO3-CIM/CR-CC(2,3) (ζ = 0.003 hartree) and MFMO3/CR-CC(2,3). The percentage of correlation (in parentheses) recovered by MFMO3-CIM/CR-CC(2,3) is relative to MFMO3/CR-CC(2,3). As $\zeta \rightarrow 0$, FMO-CIM/CR-CC(2,3) becomes equivalent to standard MFMO3/CR-CC(2,3). Similar to the ζ = 0.003 FMO2-CIM/MP2 and CIM/MP2 calculations on Ala_{20}, MFMO3-CIM/MP2 recovered ∼99.2% of the MP2 correlation of adamantane. The MFMO3-CIM/CCSD and MFMO3-CIM/CR-CC(2,3) calculations recovered ∼99.6 and ∼99.5% of the MFMO3/CCSD and MFMO3/CR-CC(2,3) correlation energies, respectively; for MFMO3-CIM/CR-CC(2,3), the correlation energy was obtained using 35% of the MFMO3/CR-CC(2,3) wall time.

IV. CONCLUSIONS

A preliminary implementation of the FMO-CIM approach to the calculation of correlation energies on large systems has been coded into GAMESS. The linear scaling, local correlation CIM algorithm is effectively extended to larger systems by replacement of the prerequisite RHF reference calculation with the family of near-linear scaling FMO methods. The natural parallelism of both FMO and CIM is retained in the

DOI: 10.1021/jp509266g
J. Phys. Chem. A 2015, 119, 3587−3593

combined method, providing efficient GDDI parallelization at the FMO fragment and CIM subsystem calculation levels. The method is, however, lacking a parallel implementation of the CIM subsystem generation step. The serial implementation of the subsystem generation step scales linearly with system size; however, a large prefactor is applied due to the large number of subsystems generated in serial for large systems. Future work will parallelize the subsystem generation step, alleviating this bottleneck.

Despite the subsystem generation bottleneck, FMO2-CIM/MP2 calculations (ζ = 0.001) on Ala_{20} recover 99.56% of the full-system MP2 correlation energy, with the ten-fragment scheme requiring 79% less wall time than full MP2 calculations. FMO-CIM/MP2 reproduces the CIM/MP2 results for both the ten- and five-fragment schemes for both choices of the subsystem size parameter ζ. Similar results were obtained for water-solvated adamantine: The MFMO3-CIM/CR-CC(2,3) calculation recovered 99.51% of the MFMO3-CR-CC(2,3) correlation energy, requiring 65% less wall time.

AUTHOR INFORMATION

Corresponding Author

*E-mail: mark@si.msg.chem.iastate.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants from the Department of Defense Productivity, Enhancement Technology Transfer and Training (PETTT) program (support for ADF) and from a grant from the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences and Biosciences through the Ames Laboratory PCTC, Chemical Physics, and Homogeneous and Interfacial Catalysis project. The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under contract no. DE-AC02-07CH11358.

REFERENCES

(1) Sinanoglu, S. Many-Electron Theory of Atoms, Molecules and their Interactions. *Adv. Chem. Phys.* **1964**, *6*, 315−358.
(2) Nesbet, R. K. Electronic Structure in Atoms and Molecules. *Adv. Chem. Phys.* **1965**, *9*, 321−363.
(3) Pulay, P. J. Localizability of Dynamic Electron Correlation. *Chem. Phys. Lett.* **1983**, *100*, 151−154.
(4) Pulay, P.; Saebø, S. Local Correlation interaction: An Efficient Approach for Larger Molecules. *Phys. Chem. Lett.* **1985**, *113*, 13−18.
(5) Pulay, P.; Saebø, S. Orbital-Invariant Formulation and Second-Order Gradient Evaluation in Møller-Plesset Perturbation Theory. Theor. *Chim. Acta.* **1986**, *69*, 357−368.
(6) Pulay, P.; Saebø, S. Fourth-Order Møller-Plettest Perturbation Theory in the Local Correlation Treatment. I. Method. *J. Chem. Phys.* **1987**, *914*, 914−922.
(7) Pulay, P.; Saebø, S. The Local Correlation Treatment. II. Implementation and Tests. *J. Chem. Phys.* **1988**, *88*, 1884−1890.
(8) Saebø, S.; Pulay, P. Local Treatment of electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.
(9) Beran, G. J. O.; Hirata, S. Fragment and Localized Orbital Methods in Electronic Structure Theory. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7559−7561.
(10) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632−672.
(11) Meyer, W. Theory of Self-Consistent Electron Pairs. An Iterative Method for Correlated Many Electron Wavefunctions. *J. Chem. Phys.* **1976**, *64*, 2901−2907.
(12) Hampel, C.; Werner, H.-J. Local Treatment of Electron Correlation in Coupled Cluster Theory. *J. Chem. Phys.* **1996**, *104*, 6286−6297.
(13) Schütz, M.; Werner, H.-J. Low-Order Scaling Local Electron Correlation Methods. IV. Linear Scaling Local Coupled-Cluster (LCCSD). *J. Chem. Phys.* **2001**, *114*, 661−681.
(14) Schütz, M. Low-Order Scaling Local Electron Correlation Methods. III. Linear Scaling Local Perturbative Triples Correction (T). *J. Chem. Phys.* **2000**, *113*, 9986−10001.
(15) Schütz, M.; Werner, H.-J. Local Perturbative Triples Correction (T) with Linear Cost Scaling. *J. Chem. Phys. Lett.* **2000**, *318*, 370−378.
(16) Schütz, M. Low-Order Scaling Local Correlation Methods. V. Connected Triples Beyond (T): Linear Scaling local CCSDT-1b. *J. Chem. Phys.* **2002**, *116*, 8772−8785.
(17) Schütz, M. A New, Fast, Semi-Direct Implementation of Linear Scaling Local Coupled Cluster Theory. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3841−3947.
(18) Yang, J.; Chan, G. K. L.; Manby, F. R.; Schütz, M.; Werner, H.-J. The Orbital-Specific-Virtual Local Coupled Cluster Singles and Doubles Method. *J. Chem. Phys.* **2012**, *136*, 144105−14421.
(19) Flocke, N.; Bartlett, R. J. A Natural Linear Scaling Coupled-Cluster Method. *J. Chem. Phys.* **2004**, *121*, 10935−10944.
(20) Neese, F.; Hansen, A.; Liakos, D. G. Efficient and Accurate Approximations to the Local Coupled Cluster Singles and Doubles Method using a Truncated Pair Natural Orbital Basis. *J. Chem. Phys.* **2009**, *131*, 064103−064118.
(21) Kobayashi, M.; Nakai, H. Divide-and-Conquer-Based Linear-Scaling Approach for Traditional and Renormalized Coupled Cluster Methods with Single, Double and Noniterative Triple Excitations. *J. Chem. Phys.* **2009**, *131*, 144108−144117.
(22) Ziolkowski, M.; Jansik, B.; Kjaergaard, T.; Jorgensen, P. Linear Scaling Coupled Cluster Method with Correlation Energy Based Error Control. *J. Chem. Phys.* **2010**, *133*, 014107−014112.
(23) Li, S.; Li, W.; Ma, J. A Quick Estimate of the Correlation Energy for Alkanes. *Chin. J. Chem.* **2003**, *21*, 1422−1429.
(24) Li, S.; Shen, J.; Jiang, Y. An Efficient Implementation of the "Cluster-in-Molecule" Approach for Local Electron Correlation Calculations. *J. Chem. Phys.* **2006**, *125*, 074109−074119.
(25) Gau, Y.; Li, W.; Li, S. Improved Cluster-in-Molecule Local Correlation Approach for Electron Correlation Calculation of Larger Systems. *J. Phys. Chem. A* **2014**, *118*, 8996−9004.
(26) Li, W.; Piecuch, P.; Gour, J. R.; Li, S. Local Correlation Calculations using Standard and Renormalized Coupled-Cluster Approaches. *J. Chem. Phys.* **2009**, *131*, 114109−114139.
(27) Li, W.; Piecuch, P.; Wei, L. Local Correlation Calculations using Standard and Renormalized Coupled-Cluster Methods. *AIP Conf. Proc.* **2009**, *1102*, 68−113.
(28) Li, W.; Piecuch, P. Multilevel Extension of the Cluster-in-Molecule Local Correlation Methodology: Merging Coupled and Møller-Plesset Perturbation Theories. *J. Phys. Chem. A* **2010**, *114*, 6721−6727.
(29) Li, W.; Piecuch, P. Improved Design of Orbital Domains within the Cluster-in-Molecule Local Correlation Framework: Single-Environment Cluster-in-Molecule Ansatz and Its Application to Local Coupled-Cluster Approach with Singles and Doubles. *J. Phys. Chem. A* **2010**, *114*, 8644−8657.
(30) Piecuch, P.; Wloch, M. Renormalized Coupled-Cluster Methods Exploiting Left Eigenstates of the Similarity-Transformed Hamiltonian. *J. Chem. Phys.* **2005**, *123*, 224105−224115.
(31) Wloch, J.; Gour, R.; Piecuch, P. Extension of the Renormalized Coupled-Cluster Methods Exploiting Left Eigenstates of the Similarity-Transformed Hamiltonian to Open-Shell Systems: A Benchmark Study. *J. Phys. Chem. A* **2007**, *111*, 11359−11382.
(32) Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musial, M. Efficient Computer Implementation of the Renormalized Coupled-Cluster

DOI: 10.1021/jp509266g
J. Phys. Chem. A 2015, 119, 3587−3593

Methods: The R-CCSD[T], R-CCSD(T), CR-CCSD[T] and CR-CCSD(T) approaches. *Comput. Phys. Commun.* **2002**, *149*, 71−96.
(33) Schmidt, M.W.; Baldridge, K. K.; Boatz, J.A.; Elbert, S. T.; Gordon, M. S.; Jenson, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, N.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(34) Ge, Y.; Gordon, M. S.; Piecuch, P. Breaking Bonds with the Left Eigenstate Completely Renormalized Coupled-Cluster Method. *J. Chem. Phys.* **2007**, *127*, 17410−174112.
(35) Boys, S. F. Construction of Some Molecular Orbitals to be Approximately Invariant to Changes from One Molecular to Another. *Rev. Mod. Phys.* **1960**, *32*, 296−299.
(36) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment Molecular Orbital Method: An Approximate Computational Method for Large Molecules. *Chem. Phys. Lett.* **1999**, *313*, 701−706.
(37) Fedorov, D. G.; Kitaura, K. The Three-Body Fragment Molecular Orbital Method for Accurate Calculations of Large Systems. *Chem. Phys. Lett.* **2006**, *433*, 182−187.
(38) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Aklyama, Y.; Uebayasi, M.; Kitaura, K. Fragment Molecular Orbital Method: Use of Approximate Electrostatic Potential. *Chem. Phys. Lett.* **2002**, *351*, 475−480.
(39) Fedorov, D. G.; Kitaura, K. On the Accuracy of the 3-Body Fragment Molecular Orbital Method (FMO) Applied to Density Functional Theory. *Chem. Phys. Lett.* **2004**, *389*, 129−134.
(40) Fedorov, D. G.; Kitaura, K. Second Order Møller-Plesset Perturbation Theory Based upon the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2004**, *121*, 2483−2490.
(41) Fedorov, D. G.; Kitaura, K. Multiconfiguration Self-Consistent-Field Theory Based upon the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2005**, *122*, 054108−054118.
(42) Fedorov, D. G.; Kitaura, K. Coupled-Cluster Theory based upon the Fragment Molecular-Orbital Method. *J. Chem. Phys.* **2005**, *123*, 134103−134114.
(43) Fedorov, D. G.; Olsen, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. A New Hierarchical Parallelization Scheme: Generalized Distributed Data Interface (GDDI), and an Application to the Fragment Molecular Orbital Method (FMO). *J. Comput. Chem.* **2004**, *25*, 872−880.
(44) Gau, Y.; Li, W.; Li, S. Improved Cluster-in-Molecule Local Correlation Approach for Electron Correlation Calculation of Large Systems. *J. Phys. Chem. A* **2014**, *118*, 8996−9004.
(45) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry. *J. Phys. Chem. A* **2001**, *105*, 293−307.

DOI: 10.1021/jp509266g
J. Phys. Chem. A 2015, 119, 3587−3593