

Linear-Scaling Systematic Molecular Fragmentation Approach for High-Level Coupled-Cluster Methods: Coupled-Cluster Meets Macromolecules

Uğur Bozkaya* and Betül Ermiş

Department of Chemistry, Hacettepe University, Ankara 06800, Turkey

E-mail: ugur.bozkaya@hacettepe.edu.tr

*To whom correspondence should be addressed

Abstract

The coupled-cluster (CC) singles and doubles with perturbative triples [CCSD(T)] method is frequently referred to as the "gold standard" of modern computational chemistry. However, the high computational cost of CCSD(T) [$O(N^7)$], where N is the number of basis functions, limits its applications to small-sized chemical systems. To address this problem, efficient implementations of linear-scaling coupled-cluster methods, which employ the systematic molecular fragmentation (SMF) approach, are reported. In this study: (1) to achieve exact linear-scaling and to obtain a pure *ab initio* approach, we revise the handling of nonbonded interactions in the SMF approach (2) a new fragmentation algorithm, which yields smaller sized fragments; hence, better fits high-level CC methods is introduced (3) the new SMF approach is integrated with the high-level CC methods, denoted by LSSMF-CC, for the first time. Performances of the LSSMF-CC approaches, such as LSSMF-CCSD(T), are compared with their canonical versions for a set of alkane molecules, C_nH_{2n+2} ($n=6-10$), which includes 142 molecules. Our results demonstrate that the LSSMF approach introduces negligible errors compared with the canonical methods, mean absolute errors (MAEs) are between 0.20–0.59 kcal mol⁻¹ for LSSMF-CCSD(T). To further assess the accuracy of the LSSMF-CCSD(T) approach, we also consider several polyethylene (PE) models. For the PE set, the error of LSSMF-CCSD(T)/cc-pVDZ with respect to the experimental polymerization energies per unit are between 0.08–0.63 kcal mol⁻¹. To illustrate the efficiency and applicability of the LSSMF-CCSD(T) approach, we consider an alkane molecule with 10004 atoms. For this molecule, the LSSMF-CCSD(T)/cc-pVTZ energy computation on a Linux cluster with 100 nodes, 4 cores and 5 GB of memory are provided to each node, is performed just in ~ 24 hours. As far as we know, this computation is an application of the CCSD(T) method on the largest chemical system to date. Overall, we conclude that (1) the LSSMF-CCSD(T) method can be reliably used for large scale chemical systems, where the canonical methods are not computationally affordable (2) the LSSMF-CCSD(T) method is very promising for accurate computation of energies in macromolecular systems (3) we believe that our study is a significant milestone in

developing CC methods for large-scale chemical systems.

It has been demonstrated that coupled-cluster (CC) methods are accurate for the prediction of molecular properties.¹⁻⁵ The coupled-cluster singles and doubles (CCSD) method⁶ provides quite accurate results for most molecular systems at equilibrium geometries, but nevertheless a triple excitations correction is required to obtain high accuracy.⁷⁻¹³ The coupled-cluster singles and doubles with perturbative triples [CCSD(T)] method^{10,11,14} provides excellent results for a broad range of chemical systems near equilibrium geometries.^{12,15-24} Therefore, the CCSD(T) method generally referred to as the "gold standard" of computational chemistry. However, the high computational cost of CCSD(T) [$O(N^7)$], where N is the number of basis functions, limits its applications to small-sized chemical systems.

There have been many attempts to development of reduced cost electron correlation methods.²⁵⁻³⁵ Some of these studies take advantages of the locality of molecular orbitals (MO), which is based on the idea that dynamic correlation is a short-range phenomenon. The introduction of "correlation domain" concept, by Pulay and co-workers,^{25,26} stimulated local correlation approaches. Nowadays, there are many variants of local CC methods, such as projected atomic orbitals based local CC methods (PAO-LCC),^{28,29} the local pair natural orbitals (LPNOs),³²⁻³⁴ the cluster-in-molecule (CIM) approach,³⁶⁻⁴⁰ and the divide-expand-consolidate (DEC) approach.^{41,42}

Alternative and more effective approaches, compared to LCC methods, to tackle the molecular size dependence problems of electronic structure theories are the molecular fragmentation approaches (MFA). Various molecular fragmentation approaches have been suggested to overcome the steep scaling problem of electronic structure methods.⁴³⁻⁴⁶ In molecular fragmentation approaches, a molecular system is broken up into small molecular units, and energies of the fragments are combined to approximate the energy of the entire system. Although, the logic behind all fragmentation approaches is similar, the formation of fragments, as well as the combination of the fragment energies, differ significantly from method to method. Molecular fragmentation methods include molecular tailoring approach (MTA),⁴⁷⁻⁴⁹ fragment molecular orbital theory (FMO),⁵⁰⁻⁵² molecular fractionation with conjugate caps

(MFCC),^{53,54} systematic molecular fragmentation (by annihilation) [SMF(A)],^{44,55–63} combined fragmentation method (CFM),^{60,64} generalized energy-based fragmentation (GEBF),⁶⁵ kernel energy method (KEM),^{66,67} molecules-in-molecules (MIM) approach,⁶⁸ many-overlapping-body expansion (MOBE),⁶⁹ and generalized many-body expansion (GMBE).⁷⁰

In terms of accuracy and general applicability, the SMF approach appears to be very attractive. The SMF energy is a sum of two-components: bonded and nonbonded. We may also call them as covalent and noncovalent terms. The number bonded fragments scales linearly [$O(n)$], where n is the number of groups, while the number of nonbonded fragments scales quadratically [$O(n^2)$]. To reduce the high cost of nonbonded fragments, Collins introduced a cutoff distance (R_{cut}), such as 2 Å. If the distance between monomers of a nonbonded fragment is smaller than R_{cut} , then it is treated with electronic structure methods; otherwise, with a simple perturbation theory approach. For branched molecules, Collins’ algorithm yields large-sized fragments compared to the chain-like linear alkanes case, which is another difficulty. This situation especially becomes problematic for high-level CC approaches, such as CCSD(T), where the computational cost increases steeply with the molecular size.

In this research, to achieve exact linear-scaling and to obtain a pure *ab initio* approach, we completely neglect all long-range nonbonded contributions since they already approach to zero. Further, we introduce a new fragmentation algorithm for the branched molecules, which yields smaller sized fragments; hence, the new algorithm better fits high-level CC methods. The new linear-scaling SMF algorithm, denoted by LSSMF, have been coded in C++ language by present authors and integrated with the DF0CC module^{24,71–78} of the PSI4 package.⁷⁹ The newly proposed LSSMF-CC approaches, such as LSSMF-CCSD, LSSMF-CCSD(T) as well as LSSMF-MP2, are applied to a series of alkane molecules to demonstrate their efficiency and accuracy.

The correlation energy for a CC method can be expressed as follows

$$\Delta E = \langle 0 | e^{-\hat{T}} \hat{H}_N e^{\hat{T}} | 0 \rangle, \quad (1)$$

where \hat{H}_N is the normal-ordered Hamiltonian operator,^{3,80} $|0\rangle$ is the reference determinant, and \hat{T} is the cluster excitation operator. For the CCSD wave function

$$\hat{T} = \hat{T}_1 + \hat{T}_2, \quad (2)$$

where \hat{T}_1 and \hat{T}_2 are single- and double-excitation operators, respectively.

$$\hat{T}_1 = \sum_i^{occ} \sum_a^{vir} t_i^a \hat{a}^\dagger \hat{i}, \quad (3)$$

$$\hat{T}_2 = \frac{1}{4} \sum_{i,j}^{occ} \sum_{a,b}^{vir} t_{ij}^{ab} \hat{a}^\dagger \hat{b}^\dagger \hat{j} \hat{i}, \quad (4)$$

where \hat{a}^\dagger and \hat{i} are the creation and annihilation operators, and t_i^a and t_{ij}^{ab} are the single and double excitation amplitudes, respectively.

The amplitude equations can be written as

$$\langle \Phi_i^a | e^{-\hat{T}} \hat{H}_N e^{\hat{T}} | 0 \rangle = 0, \quad (5)$$

$$\langle \Phi_{ij}^{ab} | e^{-\hat{T}} \hat{H}_N e^{\hat{T}} | 0 \rangle = 0, \quad (6)$$

where $\langle \Phi_i^a |$ and $\langle \Phi_{ij}^{ab} |$ are singly- and doubly-excited Slater determinants, respectively. For details of our CCSD and CCSD(T) implementations, one may refer to our previous studies.^{24,76,77}

The SMF approach starts with the molecule M divided into different "groups". Groups are sets of atoms defined by the SMF algorithm. The basic ideas involved in the method

can be illustrated for the simplest case involving a chain-like molecule containing N groups connected by single bonds:

$$M = G_1 G_2 G_3 \dots G_N, \quad (7)$$

The target is to derive an accurate value for the total electronic energy:

$$E(M) = E(G_1 G_2 G_3 \dots G_N). \quad (8)$$

The energy of the molecule M is determined by summation of the fragment (F_n), which are defined in terms of combinations of groups, energies. The sizes of the fragments depend on the "Level" of SFM, and the fragments can overlap with each other since a group can be involved in the multiple fragments. Hence, additional fragments with negative coefficients are generated to cancel the effects of multiple counting.

The bonded energy:

$$E_b = \sum_i^{N_{frag}} f_i E(F_i), \quad (9)$$

where f_i is the integer coefficient associated with the fragment F_i .

For a model system of chain containing five groups, SMF fragmentation scheme can be expressed as,

$$\text{Level 1 : } G_1 G_2 G_3 G_4 G_5 = G_1 G_2 + G_2 G_3 + G_3 G_4 + G_4 G_5 - G_2 - G_3 - G_4, \quad (10)$$

$$\text{Level 2 : } G_1 G_2 G_3 G_4 G_5 = G_1 G_2 G_3 + G_2 G_3 G_4 + G_3 G_4 G_5 - G_2 G_3 - G_3 G_4, \quad (11)$$

$$\text{Level 3 : } G_1 G_2 G_3 G_4 G_5 = G_1 G_2 G_3 G_4 + G_2 G_3 G_4 G_5 - G_2 G_3 G_4. \quad (12)$$

Thus, the fragment sizes increase with the level used. However, the number of fragments grows linearly with the size of the system. The authors have noted that the different Levels

used in SMF are related to some older concepts used in the field of theoretical thermochemistry. For example, Level 1 reactions are known as "isodesmic reactions", Level 2 is known as homodesmotic reactions, and Level 3 is known as hyperhomodesmotic reactions.

Since the bonded energy only includes nearby interactions, one should consider the nonbonded interactions between more distant groups. The nonbonded interactions may be evaluated by the following equation:

$$E_{nb} = \sum_{i>j}^{N_{frag}} f_i f_j E[F_i \cdots F_j]_{allowed}. \quad (13)$$

The "allowed" nonbonded interactions are the interactions that are not already included in E_b .

For a chain-like $C_n H_{2n+2}$ molecule with the SMF scheme (at level 3), the bonded fragments are just butane and propane fragments. The nonbonded fragments are just methane dimers with different molecular distances. The number of bonded and nonbonded fragments are given as follows,

$$N_b = (n-3)C_4H_{10} + (n-4)C_3H_8 \sim O(n), \quad (14)$$

$$N_{nb} = \frac{1}{2}(n-4)(n-3)CH_4 \cdots CH_4 \sim O(n^2). \quad (15)$$

The number of bonded fragments scales linearly with the number of carbons, while the number of nonbonded fragments scales quadratically. However, one may consider only short-range NB fragments, their number also scales linearly with the system size.

$$N_{nb}^{short-range} \sim O(n). \quad (16)$$

Hence, we introduce a nonbonded cutoff tolerance, Δ_{nb} . If the distance between the closest

atoms of two groups is larger than Δ_{nb} , then this nonbonded fragment is disregarded. We denote this algorithm by *distance based elimination* (DBE). An alternative approach is using the ratio of distance to covalent radii (DCRR) as follows:⁸¹

$$d_{ij} = \frac{||X_i^m - X_j^n||}{r_i + r_j}, \quad (17)$$

where X_i^m denotes the Cartesian position of the atom in the fragment m and r_i denotes the covalent radius of the atom. Atomic covalent radii are obtained from Cordero et al.⁸²

Before presenting our fragmentation algorithm let us define the notation: i, j, k, l, \dots for atoms; a, b, c, d, \dots for groups; and $\mu, \nu, \lambda, \sigma, \dots$ for fragments.

- (1) Define the level of SMF and tolerances for single, double, triple bonds as well as NB cutoff: Δ_{sb} , Δ_{db} , Δ_{tb} , and Δ_{nb} .
- (2) Read molecular info: Cartesian coordinates (X , Y , and Z), number of atoms (N_{atom}), atomic masses, atomic covalent radii (r_i)
- (3) Compute inter atomic distances: R_{ij} .
- (4) Compute bond order matrix: B_{ij} .
 - If $R_{ij} < r_i + r_j + \Delta_{sb}$ then $B_{ij} = 1$.
 - If $R_{ij} < r_i + r_j - \Delta_{db}$ then $B_{ij} = 2$.
 - If $R_{ij} < r_i + r_j - \Delta_{tb}$ then $B_{ij} = 3$.
 - else $B_{ij} = 0$.
- (5) Catch the first non-hydrogen atom. The first such atom is assigned to group 1 (in fact, group 0 in C++).
- (6) Assign the remaining non-H atoms. Starting the first non-H atom make a loop over atom pairs i, j . If $B_{ij} > 1$, then assign j to the group of i -th atom, G_i . Otherwise, assign it to the next group, G_{i+1} .

- (7) Catch double/triple bonded non-H atoms in different groups and merge them.
- (8) Assign the hydrogen atoms to each group according to values of B_{ij} .
- (9) Form the group connectivity matrix: L_{ab} . If two groups are connected to each other then $L_{ab}=1$, otherwise it is equal to zero. Further, determine the bonded atoms of two connected groups: LA_{ab} .
- (10) Determine the number of caps per group.
- (11) Form bonded and nonbonded domains for each group.
- (12) Form lists of groups, and bonded and nonbonded fragments according to the SMF level.
- (13) Write PSI4 input files for groups, and bonded and nonbonded fragments.

In each final fragment, bonds that are connecting groups in the fragment to other groups that are not in the fragment are "missing". These missing bonds are replaced by bonds to hydrogen atoms. The total number of hydrogen atoms added to fragments with a sign of +1 is exactly equal to the number added to fragments with a sign of -1. The position of each H atom is taken to lie along the missing bond vector at a distance which is proportional to the expected ratio of bond lengths. That is,

$$X_H = X_i + \frac{r_i + r_H}{r_i + r_j} [X_j - X_i], \quad (18)$$

where X_i denotes the Cartesian position of the atom in the fragment and X_j denotes the Cartesian position of the atom that is not available in the fragment.

Our fragmentation algorithm is identical to the one suggested by Deev and Collins⁵⁵ for unbranched chain-like molecules. However, in the case of branched molecules we propose a new algorithm. In order to illustrate the difference between two algorithms let us consider the 2,4-dimethylheptane (24DMH) molecule (Figure 1) for which the fragmentation result at level 3 was reported by Deev and Collins.⁵⁵

In the 24DMH molecule each carbon atom defines a group, totally seven groups. Fragmentation suggested by Deev and Collins yields to the following bonded fragments at level 3:⁵⁵

$$G_1G_2G_3G_4G_5G_6G_7 = G_1G_2G_3G_4G_5 + G_1G_4G_5G_6G_7 - G_1G_4G_5, \quad (19)$$

where $G_1G_2G_3G_4G_5G_6G_7$ represents the whole molecule. In this case, fragments $G_1G_2G_3G_4G_5$ and $G_1G_4G_5G_6G_7$ are formed from the combination of 5 groups. However, in the case of an open chain analog, the fragments form from the combination of 4 groups. Hence, Deev and Collins' algorithm yields fragments at different sizes for open chain and branched molecules. In the latter case, it yields much larger fragments, which may be a problem for high-level CC methods, where the computational cost increases steeply. Therefore, one of the authors (U.B.) suggests a new fragmentation algorithm for branched molecules, in which smaller sized fragments form as in the case of open chain molecules. Our algorithm yields the following bonded fragments for the 24DMH molecule at level 3:

$$\begin{aligned} G_1G_2G_3G_4G_5G_6G_7 = & G_1G_2G_3G_4 + G_1G_2G_4G_5 + G_1G_3G_4G_5 \\ & + G_1G_4G_5G_6 + G_1G_4G_5G_7 + G_4G_5G_6G_7 \\ & - G_1G_2G_4 - G_1G_3G_4 - 3 * G_1G_4G_5 - G_4G_5G_6 \\ & - G_4G_5G_7 + G_1G_4 + G_4G_5. \end{aligned} \quad (20)$$

In the fragmentation in Eq.(20), fragments formed by the combination of 4 groups are called as the main fragments. The remaining fragments are considered for chemical balance. Hence, we may call them as *neutralizing fragments* or *renormalization terms*, reminiscent of the many-body perturbation theory. Our algorithm produces $6F_4 + 5F_3 + 2F_2$, where F_i denotes a fragment formed from i different groups, whereas Deev and Collins's algorithm produces $2F_5 + F_3$. Hence, our algorithm yields lower size fragments, while Deev and

Collins’ algorithm yields a smaller number of fragments. For high-level CC computations with large basis sets, the size of a fragment is more important than the number of additional small fragments. Moreover, a group can be as small as CH_4 and H_2O but can be as large as benzene and naphthalene. Hence, in the case of large groups, such as benzene and naphthalene, decreasing the size of the fragment from F_5 to F_4 is still very important to reduce the cost even though small basis sets are employed. Therefore, our algorithm is more efficient in terms of computational cost and better fits high-level CC methods, such as CCSD(T).

Results from the HF, MP2, CCSD, CCSD(T), LSSMF-HF, LSSMF-MP2, LSSMF-CCSD, and LSSMF-CCSD(T) methods were obtained for a set of alkanes, $\text{C}_n\text{H}_{2n+2}$ ($n=5-20$), for comparison of the absolute energies. To further illustrate the efficiency of the LSSMF-CC methods, we additionally consider polyethylene structures, $\text{C}_n\text{H}_{2n+2}$ ($n=100-3334$), for computations of polymerization energies per unit at zero Kelvin. The optimized geometries and zero-point vibrational corrections for polyethylene (PE) structures were computed with the universal force field (UFF) approach.⁸³ For the alkanes Dunning’s correlation-consistent polarized valence double, triple, and quadruple- ζ basis sets (cc-pVDZ, cc-pVTZ, and cc-pVQZ) were employed with the frozen core approximation.^{84,85} The density-fitting approach was used for LSSMF methods considered.^{24,72,76,77} For the cc-pVXZ primary basis sets, cc-pVXZ-JKFIT⁸⁶ and cc-pVXZ-RI⁸⁷ auxiliary basis sets were employed for reference and correlation energies, respectively.

Previous studies demonstrated that accuracy of level 1 and level 2 approaches are not satisfactory.^{44,60} Further, level 4 and higher levels yield high computational costs. Therefore, considering the balance of accuracy and cost, level 3 appears to be the best fit for high-level CC approaches. Hence, in this study, all LSSMF energies were computed at level 3.

To assess the accuracy of the LSSMF approach with respect to the canonical methods we consider a set of alkanes, $\text{C}_n\text{H}_{2n+2}$ ($n=6-10$). For the first step of our assessment, we choose a safe cutoff value for nonbonded interactions: $\Delta_{nb} = 10.0 \text{ \AA}$. In the next section

effects of different Δ_{nb} values will be evaluated. Mean absolute errors (MAEs) of the LSSMF-HF, LSSMF-MP2, LSSMF-CCSD, and LSSMF-CCSD(T) methods with respect to canonical methods are depicted in Figure 2. For the C_6H_{14} isomers, total energies from MP2, CCSD, CCSD(T), LSSMF-MP2, LSSMF-CCSD, and LSSMF-CCSD(T) methods, and percentages of the LSSMF energies with respect to the canonical methods are reported in Table S1. The percent coverage values are in %99.9998–%100.0002. Hence, all considered LSSMF methods cover a satisfactory portion of the total energy of the full methods. The MAE values (Figure 2) in total energies are 0.17 (LSSMF-HF), 0.21 (LSSMF-MP2), 0.19 (LSSMF-CCSD), and 0.20 [LSSMF-CCSD(T)] kcal mol⁻¹. Further, the maximum absolute errors (Δ_{max}) for total energies are 0.26 (LSSMF-HF), 0.29 (LSSMF-MP2), 0.28 (LSSMF-CCSD), and 0.28 [LSSMF-CCSD(T)] kcal mol⁻¹. Hence, considering both error measures, MAE and Δ_{max} , the results of the LSSMF methods are in very good agreement with the canonical methods.

For the C_7H_{16} isomers, total energies from MP2, CCSD, CCSD(T), LSSMF-MP2, LSSMF-CCSD, and LSSMF-CCSD(T) methods, and percentages of the LSSMF energies with respect to the canonical methods are reported in Table S2. For the correlated methods, the percent coverage values are in %99.9997–%100.0003, while that of LSSMF-HF is in %99.9992–%100.0001. Hence, all considered LSSMF methods, especially correlated methods, cover a satisfactory portion of the total energy of the full methods. The MAE values (Figure 2) in total energies are 0.49 (LSSMF-HF), 0.30 (LSSMF-MP2), 0.26 (LSSMF-CCSD), and 0.26 [LSSMF-CCSD(T)] kcal mol⁻¹. Further, the Δ_{max} values for total energies are 1.40 (LSSMF-HF), 0.55 (LSSMF-MP2), 0.49 (LSSMF-CCSD), and 0.45 [LSSMF-CCSD(T)] kcal mol⁻¹. Hence, considering both error measures, MAE and Δ_{max} , the results of the LSSMF methods are in very good agreement with the full methods. Further, these results also show that the post-HF methods yield lower errors.

For the C_8H_{18} isomers, total energies from MP2, CCSD, CCSD(T), LSSMF-MP2, LSSMF-CCSD, and LSSMF-CCSD(T) methods, and percentages of the LSSMF energies with respect to the canonical methods are reported in Table S3. For the correlated methods, the per-

cent coverage values are in %99.9997–%100.0003, while that of LSSMF-HF is in %99.9992–%100.0001. Hence, all considered LSSMF methods cover a satisfactory portion of the total energy of the full methods. The MAE values (Figure 2) in total energies are 0.58 (LSSMF-HF), 0.30 (LSSMF-MP2), 0.27 (LSSMF-CCSD), and 0.31 [LSSMF-CCSD(T)] kcal mol⁻¹. Further, the Δ_{max} values for total energies are 1.38 (LSSMF-HF), 0.61 (LSSMF-MP2), 0.58 (LSSMF-CCSD), and 0.60 [LSSMF-CCSD(T)] kcal mol⁻¹. Hence, considering both error measures, MAE and Δ_{max} , the results of the LSSMF methods are again in very good agreement with the full methods. Further, these results also demonstrate that the electron correlation methods better compensate for the fragmentation effects.

For the C₉H₂₀ isomers, total energies from MP2, CCSD, CCSD(T), LSSMF-MP2, LSSMF-CCSD, and LSSMF-CCSD(T) methods, and percentages of the LSSMF energies with respect to the canonical methods are reported in Table S4. For the correlated methods, the percent coverage values are in %99.9995–%100.0004, while that of LSSMF-HF is in %99.9991–%100.0001. Hence, all considered LSSMF methods cover a satisfactory portion of the total energy of the full methods. The MAE values (Figure 2) in total energies are 1.15 (LSSMF-HF), 0.45 (LSSMF-MP2), 0.40 (LSSMF-CCSD), and 0.39 [LSSMF-CCSD(T)] kcal mol⁻¹. Further, the Δ_{max} values for total energies are 3.56 (LSSMF-HF), 1.51 (LSSMF-MP2), 1.32 (LSSMF-CCSD), and 1.14 [LSSMF-CCSD(T)] kcal mol⁻¹. Hence, considering both error measures, MAE and Δ_{max} , the results of the correlated LSSMF methods are in very good agreement with the full methods. Further, the results of C₉H₂₀ isomers clearly show that the electron correlation methods can better tolerate fragmentation errors.

For the C₁₀H₂₂ isomers, total energies from MP2, CCSD, CCSD(T), LSSMF-MP2, LSSMF-CCSD, and LSSMF-CCSD(T) methods, and percentages of the LSSMF energies with respect to the canonical methods are reported in Tables S5 and S6. For the correlated methods, the percent coverage values are in %99.9990–%100.0005, while that of LSSMF-HF is in %99.9978–%100.0001. Hence, all considered LSSMF methods cover a satisfactory portion of the total energy of the full methods. The MAE values (Figure 2) in total energies are

1.61 (LSSMF-HF), 0.59 (LSSMF-MP2), 0.58 (LSSMF-CCSD), and 0.59 [LSSMF-CCSD(T)] kcal mol⁻¹. Further, the Δ_{max} values for total energies are 5.30 (LSSMF-HF), 2.56 (LSSMF-MP2), 2.27 (LSSMF-CCSD), and 2.01 [LSSMF-CCSD(T)] kcal mol⁻¹. Hence, considering both error measures, MAE and Δ_{max} , the results of the correlated LSSMF methods are in good agreement with the canonical methods. These results demonstrate that the high-level electron correlation methods are less prone to fragmentation errors. Considering the results obtained for the whole alkane set, one can safely rely on the LSSMF-CC methods for high-accuracy studies in large-sized chemical systems, where the canonical methods are not computationally affordable.

In the second step of our assessment of the LSSMF approaches, we investigate the effect of nonbonded cutoff tolerances on the accuracy. For this purpose we consider five isomers of C₁₀H₂₂: 2,2,3,3-tetramethylhexane (2233TMH), 2,4-dimethyl-4-ethylhexane (24DM4E), 4-isopropylheptane (4IPH), 5-methylnonane (5MN), and n-decane (decane). For these molecules, the total energies of the LSSMF-CCSD(T) approach are computed with $\Delta_{nb} = 3 - 10$ Å. The errors at each Δ_{nb} value with respect to full methods are depicted in Figure 3. Our results indicate that the maximum error is generally obtained at 3 Å, as expected, and errors are kept constant starting with 6 Å. In the case of the n-decane molecule, we obtain the lowest errors at $\Delta_{nb} = 3$ Å. The reason why the lowest error is obtained at the shortest distance is that for the n-decane molecule bonded energy closer to CCSD(T) energy compared with the total LSSMF energy, which covers %100.0005 of the CCSD(T) energy. In other words, adding more nonbonded contribution, by increasing Δ_{nb} , one obtains lower energies compared with CCSD(T). Overall, even though we use $\Delta_{nb} = 10$ Å throughout this study, a Δ_{nb} value of 6.0 Å appears to be enough for the most purposes, which corresponds to a DCRR value of ~ 4.0 .

To further increase the applicability of the LSSMF-CCSD(T) approach, we also consider frozen natural orbitals (FNOs).⁸⁸⁻⁹¹ The FNO approximation is very helpful to reduce the computational cost of CCSD(T), while it introduces negligible errors with tight enough

occupation tolerances, such as 10^{-5} . To improve the FNO-CC results, we employ the δ_{MP2} correction as suggested by DePrince and Sherrill.⁹¹ With the FNO approximation, we can consider larger basis sets for the canonical methods; hence, we employ the cc-pVTZ basis set. For the n-decane and four lowest energy isomers, we obtain MAE and Δ_{max} values of 0.74 and 1.04 kcal mol⁻¹ for the LSSMF-FNO-CCSD(T) approach. Hence, the fragmentation error is tolerable for the FNO-CCSD(T) method, as in the case of CCSD(T).

To investigate the effect of basis sets, we also carry out total energy computations for the LSSMF-FNO-CCSD(T) method with cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets for three C₇H₁₆ isomers. One of these isomers is n-heptane, and others are the lowest energy isomers: 2,2,3-trimethylbutane and 2,2-dimethylhexane. The MAE values with respect to LSSMF-FNO-CCSD(T) for different basis sets are depicted in Figure 4. The MAE values are 0.33 (cc-pVDZ), 0.38 (cc-pVTZ), and 0.45 (cc-pVQZ) kcal mol⁻¹. Even though there is a slight increase with basis set size, the errors are still at the tolerable magnitudes.

To illustrate the accuracy of the LSSMF-CCSD(T) approach we also consider the C₁₀₀H₁₀₂, C₂₀₀H₄₀₂, C₃₅₀H₇₀₂, C₇₀₀H₁₄₀₂, and C₁₀₀₀H₂₀₀₂ molecules as polyethylene (PE) models. For the C_nH_{2n+2} structures, the experimental polymerization energies per unit at 0 K were reported⁹² as -21.482, -21.675, -21.758, -21.813, -21.830, and -21.868 kcal mol⁻¹ for $n = 100, 200, 350, 700, 1000$, and ∞ , respectively. Errors of the LSSMF-CCSD(T) method with respect to the experimental values are depicted in Figure 5. The error values are 0.08 (C₁₀₀H₁₀₂), 0.39 (C₂₀₀H₄₀₂), 0.51 (C₃₅₀H₇₀₂), 0.59 (C₇₀₀H₁₄₀₂), and 0.63 (C₁₀₀₀H₂₀₀₂) kcal mol⁻¹. Even though there is a slight increase with the molecule size, the errors are still at the tolerable magnitudes. Hence, this result demonstrate that the LSSMF-CCSD(T) method can be reliably used for large scale chemical systems.

In our LSSMF implementation, we form groups, bonded, and nonbonded fragments at first; then, we write all fragment input files to disk. In the third step, we simultaneously submit all fragment jobs to our Linux clusters. Finally, we collect the energy values from output files, merge them, and compute the final LSSMF energy. Hence, our implementation

is naturally parallel. The fragment formation procedure is the fastest step (Step 1). We can form all fragments just in a few minutes owing to our efficient fragmentation algorithm. Writing fragment input files generally takes several minutes (Step 2). Hence, the cost of overall computation is depend on the cost of CC jobs (Step 3), which is depend on the number of cores that available.

To illustrate the efficiency of our fragmentation algorithm we consider a set of alkanes, which includes 10004–50012 atoms. Total wall-time (in min) for the LSSMF code (Step 1 + Step 2) for the C_nH_{2n+2} ($n=3334, 6668, 10002, 13336, 16670$) set are depicted in Figure 6. For the largest member of the alkanes set considered, $C_{16670}H_{33342}$, the total time for the LSSMF code is just 8.4 minutes on a single node (1 core) computer. Hence, our LSSMF code is very efficient to form fragments and prepare necessary input files.

To illustrate the efficiency and applicability of the LSSMF-CCSD(T) approach, we consider the $C_{3334}H_{6670}$ molecule, which includes 10004 atoms. For the $C_{3334}H_{6670}$ molecule, the LSSMF-CCSD(T)/cc-pVTZ energy computation is performed in a Linux cluster with 100 nodes, 4 cores and 5 GB of memory are provided to each node. In this system, the total wall time for energy computation is ~ 24 hours, which indicates that the introduced method is extremely efficient. As far as we know, the largest computation reported at the CCSD(T) level is the study of Nagy and Kállay,⁹³ where a protein molecule with 1023 atoms is studied at the LNO-CCSD(T)/def2-QZVP level, which includes 44712 basis functions. In this research, we report an alkane molecule with 10004 atoms at the LSSMF-CCSD(T)/cc-pVTZ, which includes 193400 basis functions. Hence, we believe that our study is a significant milestone in developing CC methods for large-scale chemical systems.

In this research, efficient implementations of linear-scaling coupled-cluster methods, which employ the systematic molecular fragmentation approach, have been reported. For the branched molecules a new fragmentation algorithm, which yields smaller sized fragments compare with previous studies, has been introduced. The new linear-scaling SMF algorithm is denoted by LSSMF. Performances of the developed LSSMF-CC approaches, such

as LSSMF-CCSD and LSSMF-CCSD(T), have been compared with their canonical versions for a set of alkane molecules, C_nH_{2n+2} ($n=6-10$), which includes 142 molecules. Our results demonstrate that the LSSMF approach introduces negligible errors compared with the canonical methods. For the alkanes set, the MAE values are between 0.19–0.58 and 0.20–0.59 kcal mol⁻¹ for the LSSMF-CCSD and LSSMF-CCSD(T) methods, respectively. A similar performance has been observed in the case of frozen natural orbitals based CCSD(T) approach [LSSMF-FNO-CCSD(T)]. Further, we investigate basis set effects on the LSSMF methods using cc-pVXZ (X=D,T,Q) basis sets. Our results indicate that the performance of LSSMF-FNO-CCSD(T) approach with large basis sets is similar to the small basis set cases.

To further assess the accuracy of the LSSMF-CCSD(T) approach, we also consider the $C_{100}H_{102}$, $C_{200}H_{402}$, $C_{350}H_{702}$, $C_{700}H_{1402}$, and $C_{1000}H_{2002}$ molecules as polyethylene (PE) models. For the PE set, polymerization energies per unit (at 0 K) at LSSMF-CCSD(T)/cc-pVDZ level have been computed and compared with the experimental values. For the PE set, the error of LSSMF-CCSD(T) with respect to the experimental values are between 0.08–0.63 kcal mol⁻¹. Hence, our results show that the LSSMF-CCSD(T) method can be reliably used for large scale chemical systems, where the canonical methods are not computationally affordable.

To illustrate the efficiency and applicability of the LSSMF-CCSD(T) approach, we consider an alkane molecule with 10004 atoms. For the $C_{3334}H_{6670}$ molecule, the LSSMF-CCSD(T)/cc-pVTZ energy computation on a Linux cluster with 100 nodes, 4 cores and 5 GB of memory are provided to each node, is performed just in ~ 24 hours. Our results demonstrate that the the LSSMF-CCSD(T) method is extremely efficient. As far as we know, this computation is an application of the CCSD(T) method on the largest chemical system to date. Hence, we believe that our study is a significant milestone in developing CC methods for large-scale chemical systems. Overall, we conclude that the LSSMF-CCSD(T) method is very promising for accurate computation of molecular properties in macromolecular systems.

Supporting Information Available

- (1) Optimized geometries of species considered.
- (2) Absolute energies for chemical systems considered.

This material is available free of charge via the Internet at <http://pubs.acs.org/>.

Acknowledgement

This research was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK-119C023).

References

- (1) Bartlett, R. J. Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359–401.
- (2) Bartlett, R. J. Coupled-Cluster Approach to Molecular Structure and Spectra: A Step Toward Predictive Quantum Chemistry. *J. Phys. Chem.* **1989**, *93*, 1697–1708.
- (3) Crawford, T. D.; Schaefer, H. F. An Introduction to Coupled Cluster Theory for Computational Chemists. *Rev. Comp. Chem.* **2000**, *14*, 33–136.
- (4) Bartlett, R. J.; Musiał, M. Coupled-Cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291–352.
- (5) Bartlett, R. J. Coupled-Cluster Theory and Its Equation-of-Motion Extensions. *WIREs Comput. Mol. Sci.* **2012**, *2*, 126–138.
- (6) Purvis, G. D.; Bartlett, R. J. A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.* **1982**, *76*, 1910–1918.

- (7) Bartlett, R. J.; Sekino, H.; Purvis, G. D. Comparison of MBPT and Coupled-cluster Methods with Full CI. Importance of Triplet Excitation and Infinite Summations. *Chem. Phys. Lett.* **1983**, *98*, 66–71.
- (8) Lee, Y. S.; Kucharski, S. A.; Bartlett, R. J. A Coupled Cluster Approach with Triple Excitations. *J. Chem. Phys.* **1984**, *81*, 5906–5912.
- (9) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. Quadratic Configuration Interaction. A General Technique for Determining Electron Correlation Energies. *J. Chem. Phys.* **1987**, *87*, 5968–5975.
- (10) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electronic Correlation Theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- (11) Bartlett, R. J.; Watts, J. D.; Kucharski, S. A.; Noga, J. Non-Iterative Fifth-Order Triple and Quadruple Excitation Energy Corrections in Correlated Methods. *Chem. Phys. Lett.* **1990**, *165*, 513–522.
- (12) Scuseria, G. E.; Lee, T. J. Comparison of Coupled-Cluster Methods which Include the Effects of Connected Triple Excitations. *J. Chem. Phys.* **1990**, *93*, 5851–5855.
- (13) Scuseria, G. E.; Hamilton, T. P.; Schaefer, H. F. An Assessment for the Full Coupled Cluster Method Including All Single, Double, and Triple Excitations: The Diatomic Molecules LiH, Li₂, BH, LiF, C₂, BeO, CN⁺, BF, NO⁺, and F₂. *J. Chem. Phys.* **1990**, *92*, 568–573.
- (14) Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. Towards a Full CCSDT Model for Electron Correlation. *J. Chem. Phys.* **1985**, *83*, 4041–4046.
- (15) Lee, T. J.; Scuseria, G. E. In *Quantum Mechanical Electronic Structure Calculations*

- with Chemical Accuracy*; Langhoff, S. R., Ed.; Kluwer Academic: Dordrecht, 1995; pp 47–108.
- (16) Watts, J. D.; Stanton, J. F.; Bartlett, R. J. A Benchmark Coupled-Cluster Single, Double, and Triple Excitation (CCSDT) Study of the Structure and Harmonic Vibrational Frequencies of the Ozone Molecule. *Chem. Phys. Lett.* **1991**, *178*, 471–474.
- (17) Scuseria, G. E. Ab Initio Theoretical Predictions of the Equilibrium Geometries of C_{60} , $C_{60}H_{60}$ and $C_{60}F_{60}$. *Chem. Phys. Lett.* **1991**, *176*, 423–427.
- (18) Gauss, J.; Lauderdale, W. J.; Stanton, J. F.; Watts, J. D.; Bartlett, R. J. Analytic Energy Gradients for Open-Shell Coupled-Cluster Singles and Doubles (CCSD) Calculations Using Restricted Open-Shell Hartree–Fock (ROHF) Reference Functions. *Chem. Phys. Lett.* **1991**, *182*, 207–215.
- (19) Watts, J. D.; Bartlett, R. J. Coupled-Cluster Calculations on the C_2 molecule and the C_2^+ and C_2^- Molecular Ions. *J. Chem. Phys.* **1992**, *96*, 6073–6084.
- (20) Thomas, J. R.; DeLeeuw, B. J.; Vacek, G.; Crawford, T. D.; Yamaguchi, Y.; Schaefer, H. F. The Balance Between Theoretical Method and Basis Set Quality: A Systematic Study of Equilibrium Geometries, Dipole Moments, Harmonic Vibrational Frequencies, and Infrared Intensities. *J. Chem. Phys.* **1993**, *99*, 403–416.
- (21) Watts, J. D.; Gauss, J.; Bartlett, R. J. Coupled-Cluster Methods With Noniterative Triple Excitations For Restricted Open-Shell Hartree–Fock And Other General Single Determinant Reference Functions. Energies And Analytical Gradients. *J. Chem. Phys.* **1993**, *98*, 8718–8733.
- (22) Crawford, T. D.; Schaefer, H. F. A Comparison of Two Approaches to Perturbational Triples Corrections to the Coupled-Cluster Singles and Doubles Method for High-Spin Open-Shell Systems. *J. Chem. Phys.* **1996**, *104*, 6259–6264.

- (23) Crawford, T. D.; Lee, T. J.; Schaefer, H. F. A New Spin-Restricted Perturbative Triple Excitation Correction for Coupled Cluster Theory. *J. Chem. Phys.* **1997**, *107*, 7943–7950.
- (24) Bozkaya, U.; Sherrill, C. D. Analytic Energy Gradients for The Coupled-Cluster Singles and Doubles with Perturbative Triples Method with the Density-Fitting Approximation. *J. Chem. Phys.* **2017**, *147*, 044104–044114.
- (25) Saebo, S.; Tong, W.; Pulay, P. Efficient Elimination of Basis Set Superposition Errors by the Local Correlation Method: Accurate Ab Initio Studies of The Water Dimer. *J. Chem. Phys.* **1993**, *98*, 2170–2175.
- (26) Saebo, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213–236.
- (27) Schutz, M.; Hetzer, G.; Werner, H.-J. Low-Order Scaling Local Electron Correlation Methods. I. Linear Scaling Local MP2. *J. Chem. Phys.* **1999**, *111*, 5691–5705.
- (28) Hampel, C.; Werner, H.-J. Local Treatment of Electron Correlation in Coupled Cluster Theory. *J. Chem. Phys.* **1996**, *104*, 6286–6297.
- (29) Schutz, 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.
- (30) Rauhut, G.; Pulay, P.; Werner, H.-J. Integral Transformation with Low-Order Scaling for Large Local Second-Order Moller-Plesset Calculations. *J. Comput. Chem.* **1998**, *19*, 1241–1254.
- (31) Schutz, M.; Werner, H.-J. Local Perturbative Triples Correction (T) with Linear Cost Scaling. *Chem. Phys. Lett.* **2000**, *318*, 370–378.
- (32) Neese, F.; Hansen, A.; Liakos, D. G. Efficient and Accurate Approximations to the Local

- Coupled Cluster Singles Doubles Method Using a Truncated Pair Natural Orbital Basis. *J. Chem. Phys.* **2009**, *131*, 064103.
- (33) Riplinger, C.; Neese, F. An Efficient and Near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *J. Chem. Phys.* **2013**, *138*, 034106.
- (34) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural Triple Excitations in Local Coupled Cluster Calculations with Pair Natural Orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.
- (35) Rolik, Z.; Szegedy, L.; Ladjánszki, I.; Ladóczki, B.; Kállay, M. An Efficient Linear-scaling Ccsd(T) Method Based on Local Natural Orbitals. *J. Chem. Phys.* **2013**, *139*, 094105.
- (36) Li, S.; Ma, J.; Jiang, Y. Linear Scaling Local Correlation Approach for Solving The Coupled Cluster Equations of Large Systems. *J. Comp. Chem.* **2001**, *23*, 237–244.
- (37) 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.
- (38) 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.
- (39) Li, W.; Piecuch, P. Multilevel Extension of the Cluster-in-Molecule Local Correlation Methodology: Merging Coupled-Cluster and Møller-Plesset Perturbation Theories. *J. Phys. Chem. A* **2010**, *114*, 6721–6727.
- (40) Rolik, Z.; Kállay, M. A General-Order Local Coupled-Cluster Method Based on the Cluster-in-Molecule Approach. *J. Chem. Phys.* **2011**, *135*, 104111.

- (41) Ziółkowski, M.; Jansík, B.; Kjærgaard, T.; Jørgensen, P. Linear Scaling Coupled Cluster Method with Correlation Energy Based Error Control. *J. Chem. Phys.* **2010**, *133*, 014107.
- (42) Eriksen, J. J.; Baudin, P.; Ettenhuber, P.; Kristensen, K.; Kjærgaard, T.; Jørgensen, P. Linear-Scaling Coupled Cluster with Perturbative Triple Excitations: The Divide–Expand–Consolidate CCSD(T) Model. *J. Chem. Theory Comput.* **2015**, *11*, 2984–2993.
- (43) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2011**, *112*, 632–672.
- (44) Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.* **2015**, *115*, 5607–5642.
- (45) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643–5677.
- (46) Akimov, A. V.; Prezhdo, O. V. Large-Scale Computations in Chemistry: A Bird’s Eye View of a Vibrant Field. *Chem. Rev.* **2015**, *115*, 5797–5890.
- (47) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. Molecular Tailoring Approach for Simulation of Electrostatic Properties. *J. Phys. Chem.* **1994**, *98*, 9165–9169.
- (48) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. Molecular Tailoring Approach for Geometry Optimization of Large Molecules: Energy Evaluation and Parallelization Strategies. *J. Chem. Phys.* **2006**, *125*, 104109.
- (49) Sahu, N.; Gadre, S. R. Molecular Tailoring Approach: A Route for ab Initio Treatment of Large Clusters. *Acc. Chem. Res.* **2014**, *47*, 2739–2747.
- (50) 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.
- (51) Mochizuki, Y.; Koikegami, S.; Nakano, T.; Amari, S.; Kitaura, K. Large Scale MP2 Calculations with Fragment Molecular Orbital Scheme. *Chem. Phys. Lett.* **2004**, *396*, 473–479.
- (52) Mochizuki, Y.; Yamashita, K.; Nakano, T.; Okiyama, Y.; Fukuzawa, K.; Taguchi, N.; Tanaka, S. Higher-Order Correlated Calculations Based on Fragment Molecular Orbital Scheme. *Theor. Chem. Acc.* **2011**, *130*, 515–530.
- (53) Zhang, D. W.; Zhang, J. Z. H. Molecular Fractionation with Conjugate Caps for Full Quantum Mechanical Calculation of Protein–Molecule Interaction Energy. *J. Chem. Phys.* **2003**, *119*, 3599–3605.
- (54) He, X.; Zhu, T.; Wang, X.; Liu, J.; Zhang, J. Z. H. Fragment Quantum Mechanical Calculation of Proteins and Its Applications. *Acc. Chem. Res.* **2014**, *47*, 2748–2757.
- (55) Deev, V.; Collins, M. A. Approximate Ab Initio Energies by Systematic Molecular Fragmentation. *J. Chem. Phys.* **2005**, *122*, 154102.
- (56) Collins, M. A.; Deev, V. A. Accuracy and Efficiency of Electronic Energies from Systematic Molecular Fragmentation. *J. Chem. Phys.* **2006**, *125*, 104104.
- (57) Collins, M. A. Ab Initio Lattice Dynamics of Nonconducting Crystals by Systematic Fragmentation. *J. Chem. Phys.* **2011**, *134*, 164110.
- (58) Frankcombe, T. J.; Collins, M. A. Potential Energy Surfaces For Gas-surface Reactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8379–8391.
- (59) Collins, M. A. Systematic Fragmentation of Large Molecules By Annihilation. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7744–7751.

- (60) Collins, M. A.; Cvitkovic, M. W.; Bettens, R. P. A. The Combined Fragmentation and Systematic Molecular Fragmentation Methods. *Acc. Chem. Res.* **2014**, *47*, 2776–2785.
- (61) Collins, M. A. Molecular Forces, Geometries, and Frequencies by Systematic Molecular Fragmentation Including Embedded Charges. *J. Chem. Phys.* **2014**, *141*, 094108.
- (62) Kobayashi, R.; Amos, R.; Collins, M. A. Microsolvation within the Systematic Molecular Fragmentation by Annihilation Approach. *J. Phys. Chem. A* **2017**, *121*, 334–341.
- (63) Kobayashi, R.; Amos, R. D.; Reid, D. M.; Collins, M. A. Application of the Systematic Molecular Fragmentation by Annihilation Method to ab Initio NMR Chemical Shift Calculations. *J. Phys. Chem. A* **2018**, *122*, 9135–9141.
- (64) Le, H.-A.; Tan, H.-J.; Ouyang, J. F.; Bettens, R. P. A. Combined Fragmentation Method: A Simple Method for Fragmentation of Large Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 469–478.
- (65) Li, S.; Li, W.; Ma, J. Generalized Energy-Based Fragmentation Approach and Its Applications to Macromolecules and Molecular Aggregates. *Acc. Chem. Res.* **2014**, *47*, 2712–2720.
- (66) Huang, L.; Massa, L.; Karle, J. Kernel Energy Method Illustrated with Peptides. *Int. J. Quant. Chem.* **2005**, *103*, 808–817.
- (67) Huang, L.; Massa, L.; Karle, J. Kernel energy method: Application to insulin. *Proc. Natl. Acad. Sci* **2005**, *102*, 12690–12693.
- (68) Mayhall, N. J.; Raghavachari, K. Molecules-in-Molecules: An Extrapolated Fragment-Based Approach for Accurate Calculations on Large Molecules and Materials. *J. Chem. Theory Comput.* **2011**, *7*, 1336–1343.

- (69) Mayhall, N. J.; Raghavachari, K. Many-Overlapping-Body (MOB) Expansion: A Generalized Many Body Expansion for Nondisjoint Monomers in Molecular Fragmentation Calculations of Covalent Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 2669–2675.
- (70) Richard, R. M.; Herbert, J. M. A Generalized Many-body Expansion and a Unified View of Fragment-based Methods In Electronic Structure Theory. *J. Chem. Phys.* **2012**, *137*, 064113.
- (71) Bozkaya, U. Orbital-Optimized Second-Order Perturbation Theory with Density-Fitting and Cholesky Decomposition Approximations: An Efficient Implementation. *J. Chem. Theory Comput.* **2014**, *10*, 2371–2378.
- (72) Bozkaya, U. Derivation of General Analytic Gradient Expressions for Density-Fitted Post-Hartree-Fock Methods: An Efficient Implementation for the Density-Fitted Second-Order Møller–Plesset Perturbation Theory. *J. Chem. Phys.* **2014**, *141*, 124108.
- (73) Bozkaya, U. Analytic Energy Gradients and Spin Multiplicities for Orbital-Optimized Second-Order Perturbation Theory with Density-Fitting Approximation: An Efficient Implementation. *J. Chem. Theory Comput.* **2014**, *10*, 4389–4399.
- (74) Bozkaya, U. Orbital-Optimized MP3 and MP2.5 with Density-Fitting and Cholesky Decomposition Approximations. *J. Chem. Theory Comput.* **2016**, *12*, 1179–1188.
- (75) Bozkaya, U. Orbital-Optimized Linearized Coupled-Cluster Doubles with Density-Fitting and Cholesky Decomposition Approximations: An Efficient Implementation. *Phys. Chem. Chem. Phys.* **2016**, *18*, 11362–11373.
- (76) Bozkaya, U.; Sherrill, C. D. Analytic Energy Gradients for the Coupled-Cluster Singles and Doubles Method with the Density-Fitting Approximation. *J. Chem. Phys.* **2016**, *144*, 174103.

- (77) Bozkaya, U. A Noniterative Asymmetric Triple Excitation Correction for The Density-Fitted Coupled-Cluster Singles and Doubles Method: Preliminary Applications. *J. Chem. Phys.* **2016**, *144*, 144108.
- (78) Bozkaya, U. Analytic Energy Gradients for Orbital-Optimized MP3 and MP2.5 with the Density-Fitting Approximation: An Efficient Implementation. *J. Comp. Chem.* **2018**, *39*, 351–360.
- (79) Smith, D. G. A.; Burns, L. A.; Simmonett, A. C.; Parrish, R. M.; Schieber, M. C.; Galvelis, R.; Kraus, P.; Kruse, H.; Remigio, R. D.; Alenaizan, A.; James, A. M.; Lehtola, S.; Misiewicz, J. P.; Scheurer, M.; Shaw, R. A.; Schriber, J. B.; Xie, Y.; Glick, Z. L.; Sirianni, D. A.; O’Brien, J. S.; Waldrop, J. M.; Kumar, A.; Hohenstein, E. G.; Pritchard, B. P.; Brooks, B. R.; Schaefer, H. F.; Sokolov, A. Y.; Patkowski, K.; DePrince, A. E.; Bozkaya, U.; King, R. A.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. Psi4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **2020**, *152*, 184108.
- (80) Shavitt, I.; Bartlett, R. J. *Many-Body Methods in Chemistry and Physics*, 1st ed.; Cambridge Press: New York, 2009; pp 443–449.
- (81) Addicoat, M. A.; Collins, M. A. Accurate Treatment of Nonbonded Interactions within Systematic Molecular Fragmentation. *J. Chem. Phys.* **2009**, *131*, 104103.
- (82) Cordero, B.; Gómez, V.; Platero-Prats, A. E.; Revés, M.; Echeverría, J.; Cremades, E.; Barragán, F.; Alvarez, S. Covalent Radii Revisited. *Dalton Trans.* **2008**, 2832–2838.
- (83) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (84) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

- (85) Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. V. Core-Valence Basis Sets for Boron through Neon. *J. Chem. Phys.* **1995**, *103*, 4572–4585.
- (86) Weigend, F. A fully Direct RI-HF Algorithm: Implementation, Optimised Auxiliary Basis Sets, Demonstration of Accuracy and Efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.
- (87) Weigend, F.; Köhn, A.; Hättig, C. Efficient Use of the Correlation Consistent Basis Sets in Resolution of the Identity MP2 Calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- (88) Taube, A. G.; Bartlett, R. J. Frozen Natural Orbitals: Systematic Basis Set Truncation for Coupled-Cluster Theory. *Collect. Czech. Chem. Commun.* **2005**, *70*, 837–850.
- (89) Taube, A. G.; Bartlett, R. J. Frozen Natural Orbital Coupled-Cluster Theory: Forces and Application to Decomposition of Nitroethane. *J. Chem. Phys.* **2008**, *128*, 164101.
- (90) Landau, A.; Khistyayev, K.; Dolgikh, S.; Krylov, A. I. Frozen Natural Orbitals for Ionized States within Equation-of-Motion Coupled-Cluster Formalism. *J. Chem. Phys.* **2010**, *132*, 014109.
- (91) DePrince, A. E.; Sherrill, C. D. Accurate Noncovalent Interaction Energies Using Truncated Basis Sets Based on Frozen Natural Orbitals. *J. Chem. Theory Comput.* **2013**, *9*, 293–299.
- (92) Jessup, R. S. The Heat and Free Energy of Polymerization of Ethylene. *J. Chem. Phys.* **1948**, *16*, 661–664.
- (93) Nagy, P. R.; Kállay, M. Approaching the Basis Set Limit of CCSD(T) Energies for Large Molecules with Local Natural Orbital Coupled-Cluster Methods. *J. Chem. Theory Comput.* **2019**, *15*, 5275–5298.

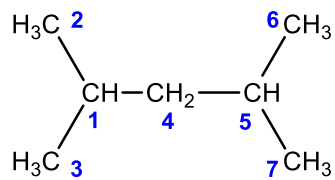


Figure 1: 2,4-dimethylheptane (24DMH) molecule.

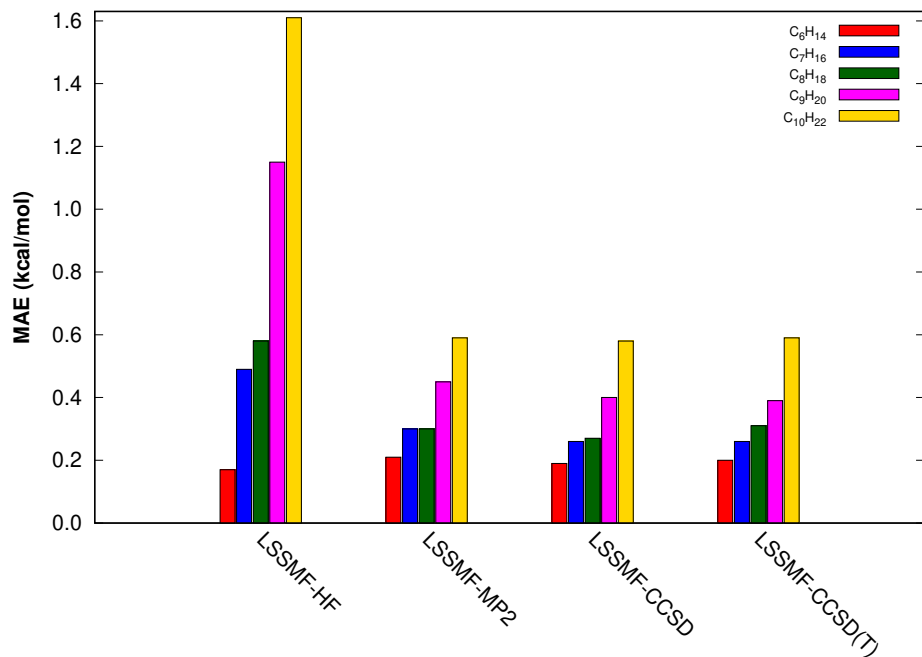


Figure 2: Mean absolute errors in the total energies of the C_nH_{2n+2} ($n=6-10$) isomers for the LSSMF-HF, LSSMF-MP2, LSSMF-CCSD, and LSSMF-CCSD(T) methods with respect to canonical methods. All computations are performed with the cc-pVDZ basis set and with the $\Delta_{nb} = 10.0$ Å.

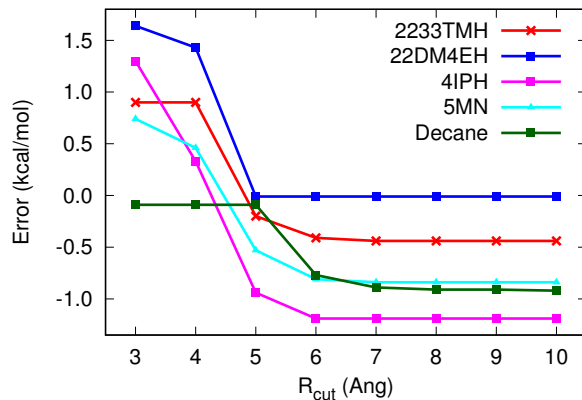


Figure 3: Errors of the LSSMF-CCSD(T) method with respect to the full method with different cutoff distances for 2,2,3,3-tetramethylhexane (2233TMH), 2,4-dimethyl-4-ethylhexane (24DM4E), 4-isopropylheptane (4IPH), 5-methylnonane (5MN), and n-decane (Decane) molecules. All computations are performed with the cc-pVDZ basis set and with $\Delta_{nb} = 10.0$ Å.

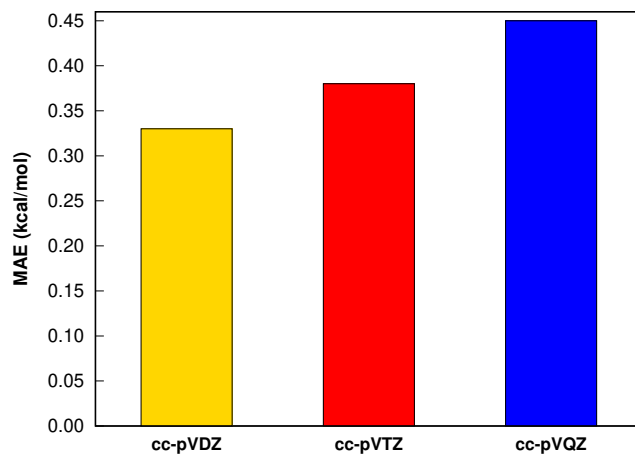


Figure 4: Mean absolute errors in the total energies of three C_7H_6 isomers for the LSSMF-FNO-CCSD(T) method with respect to FNO-CCSD(T). All computations are performed with the FNO occupation tolerance of 10^{-5} and $\Delta_{nb} = 10.0$ Å in the cc-pVXZ (X=D,T,Q) basis sets.

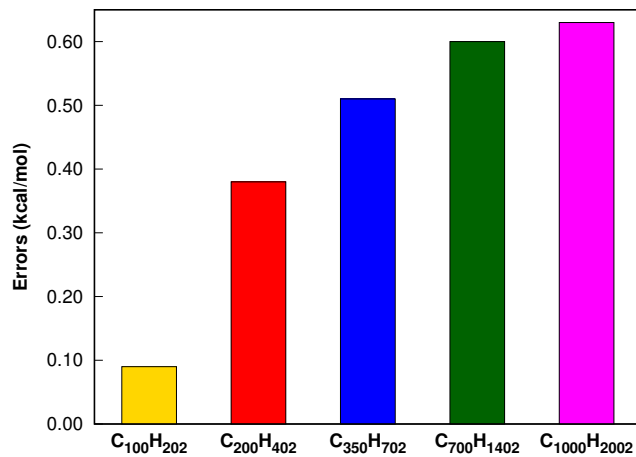


Figure 5: Errors in the polymerization energy per unit (at 0 K) for the LSSMF-CCSD(T) method with respect to the experimental values for the $C_{100}H_{102}$, $C_{200}H_{402}$, $C_{350}H_{702}$, $C_{700}H_{1402}$, and $C_{1000}H_{2002}$ molecules. All computations are performed with the $\Delta_{nb} = 10.0$ Å in a cc-pVDZ basis set.

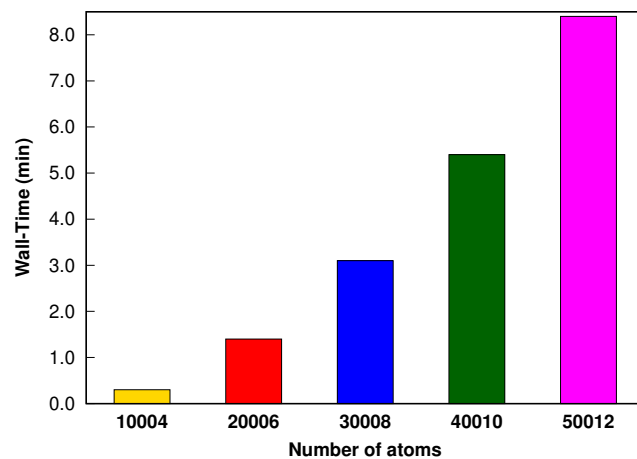


Figure 6: Total wall-time (in min) for the LSSMF code for a C_nH_{2n+2} set. All procedures were performed on a single node (1 core) Intel(R) Xeon(R) Gold 5218 CPU @ 2.30 GHz computer.

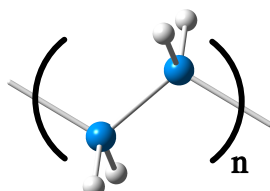
TOC Graphic

LSSMF-CCSD(T)/cc-pVTZ

193400 basis functions

100 nodes, 4 cores per node

~24 hours



$\text{C}_{3334}\text{H}_{6670}$