JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

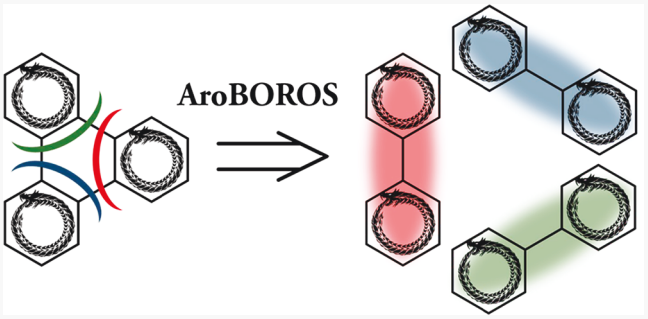
Aromatic Fragmentation Based on a Ring Overlap Scheme: An Algorithm for Large Polycyclic Aromatic Hydrocarbons Using the Molecules-in-Molecules Fragmentation-Based Method

Benjamin W. Noffke,[†] Daniel Beckett,[†] Liang-shi Li, and Krishnan Raghavachari*

Cite This: *J. Chem. Theory Comput.* 2020, 16, 2160−2171 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We present a novel and systematic fragmentation scheme to treat polycyclic aromatic hydrocarbons (PAHs) built off the molecules-in-molecules composite method. Our algorithm generates a set of biphenyl and naphthalene subsystems overlapping by whole sextet rings, ensuring all calculations are performed on aromatic molecules. Hence, our method is called Aromatic Fragmentation Based on a Ring Overlap Scheme (AroBOROS), and the generated fragments may be combined to form a hierarchy of subsystems to reduce errors for more complex PAHs. Errors are reduced to below chemical accuracy by combining subsystems that reflect the lowest energy structures determined by Clar's rule of aromatic sextets, and this is shown on two diverse test sets of PAHs ranging from 18 to 84 carbon atoms. Additionally, evaluations are performed for larger PAHs, as well as a nanotube fragment, containing up to 132 carbon atoms, and it is shown that good results may be achieved even with fragments representing an appreciably small portion of the full system.



I. INTRODUCTION

Electronic structure theory has long been able to reproduce the energies of small systems within chemical accuracy through the application of sophisticated theoretical methods, such as those based on coupled cluster theory, but with the caveat of steep computational scaling.[1−16] However, as the field grows so does the average size of the systems of interest, and even DFT with appropriately sized basis sets can be computationally demanding or intractable for sufficiently large molecules. While efforts are ongoing in accelerating the formally $O(N^7)$ scaling CCSD(T) and related methods,[17,18] the recent emergence of fragmentation-based composite methods has led to an array of techniques able to approximate computationally expensive levels of theory in a way that scales asymptotically linearly with the size of the system. Additionally, the majority of advances made in reducing the scaling of correlated methods or in adapting to hardware advances, such as massively parallel computer architectures,[19,20] can be taken into account in a fragmentation method most efficiently. The versatility of these types of methods implies they will remain a frontline of exploration for some time, and it is in our best interest to expand the variety of chemical systems fragmentation methods can be applied to.

In general, fragmentation methods divide a calculation on a large system into many small calculations on specially selected subsystems (fragments) which are assembled to extrapolate the energy of the whole molecule to the target level of theory.[21,22] In our recent review, a useful binary classification has been put forward sorting fragmentation methods as either "top-down" or "bottom-up", depending on the way fragments are constructed.[22,23] Top-down approaches create overlapping subsystems and follow a scheme based on the inclusion−exclusion principle to handle the overlapping regions that are overcounted when summing up individual subsystems. Examples of top-down methods relying on overlapping subsystems include the molecular tailoring approach (MTA, notably the first method to apply the inclusion−exclusion principle to fragmentation techniques),[24−27] generalized energy based fragmentation (GEBF),[28] molecules-in-molecules (MIM),[29] molecular fractionation with conjugate caps (MFCC),[30] systematic molecular fragmentation (SMF),[31] and the combined fragment-based method (CFM).[32] Bottom-up methods, on the other hand, divide a system into nonoverlapping monomer fragments which can be combined to form *n*-body subsystems treated with the many-body expansion in the basis of the monomers. Such methods include the fragment molecular orbital method (FMO),[33] kernel energy method (KEM),[34] electrostatically embedded many body method (EE-MB),[35] hybrid many body

Received: June 7, 2019
Published: March 2, 2020

interaction (HMBI),[36] multilevel fragment-based approach (MFBA),[37] and generalized many body expansion (GMBE).[38] Recent and comprehensive reviews of fragmentation methods can be found in refs 21 and 22.

While fragmentation methods have been expanded and explored extensively in the last several years, these techniques have largely been limited to nonconjugated molecules. A major consideration in fragmentation-based methods, and composite methods in general, is how to satisfy the valences of a cut bond in the calculation of a model subsystem. Generally, a hydrogen link atom is placed along the vector of the broken bond, and this is only valid when cutting a single bond.[39] Given this, the majority of fragmentation methods begin by identifying all single bonds in the system and then deciding where to cut to form fragments.

In polycyclic aromatic hydrocarbons (PAHs), each carbon−carbon bond possesses a bond order greater than one, making it impossible to avoid cutting a "multiple" bond. Previous approaches for fragmenting large PAHs have been implemented in the MTA,[40] KEM,[41] and MFCC[42] frameworks and rely on generating large fragments to circumvent the effects of disturbing aromaticity. Both MTA and MFCC are top-down approaches producing subsystems that overlap by whole ring structures; however, neither of these is a systematic approach able to be applied "black box" to a large class of PAHs. In order to decrease the size of the fragment subsystems and motivate a hierarchy of potential primary subsystems, herein we develop a method of systematically creating fragments in a PAH by taking into account the nature of aromatic systems and only choosing the least detrimental bonds to cut. We call our algorithm Aromatic Fragmentation Based on a Ring Overlap Scheme (AroBOROS), and an explanation for the acronym is discussed in the Supporting Information (SI). Furthermore, in the Results and Discussion section, we compare AroBOROS to the three approaches that have been used previously to illustrate the importance of a systematic hierarchy keeping fragments small.

When considering the aromaticity of a PAH, one can determine the most contributing resonance structures by following Clar's rule of sextets.[43] Clar's rule prioritizes maximizing the number of rings containing six pi electrons, forming an aromatic sextet. It has been experimentally verified that rings with two and four pi electrons in the dominant Clar structure are prone to reactive behavior closer to alkenes under similar conditions.[43,44] PAHs with only aromatic sextets, all other rings being "empty rings" devoid of pi electrons, tend to be the most stable and are referred to as fully benzenoid. This valence bond treatment of pi electrons even manifests itself physically; in an atomic force microscopy investigation of the molecule hexabenzocoronene (HBC), the sextet rings have been observed to have shorter carbon−carbon bond lengths than the empty rings.[45] In HBC, and PAHs in general, not all aromatic bonds are equal and the bonds connecting sextet rings to other sextet rings have more single bond character than those bonds forming a sextet ring. From this, HBC can be considered as seven embedded benzene rings. Creating another representation, these rings might be considered irreducible "superatom" units forming the full system. The purpose of understanding the Clar structures is to interpret the subsystems generated by AroBOROS and to justify the choice of subsystems in the fragmentation scheme.

AroBOROS is largely formulated for use with the molecules-in-molecules (MIM) treatment, generating special primary subsystems that are passed to MIM. MIM is an energy-based, top-down fragmentation method that employs overlapping fragments referred to as primary subsystems. MIM has been applied extensively to biomolecules for the evaluation of their energies as well as their molecular properties.[46] In addition to being a fragmentation method, MIM can use two (or more) layers, similar to the hybrid method ONIOM,[47] to combine two (or more) different levels of theory and vastly improve its energy extrapolation.[48] Interfacing with MIM also means the current formulation of the algorithm in this paper can be used for geometry optimization and to obtain all the spectroscopic properties MIM has been extended to calculate, including IR, Raman, VCD, ROA, and NMR,[49−53] though we focus on absolute energies here as a test to ensure our fragmentation scheme is accurate.

The first step in MIM is to identify the primitive fragments of a system, which are generated by cutting all single bonds between non-hydrogen atoms in the system. These primitive fragments are combined following a scheme based on a distance-, number-, or connectivity-based criteria to form overlapping primary subsystems. The inclusion−exclusion principle (IEP) is used to compensate for the overcounting that results from the overlapping regions of the primary subsystems. Applying the IEP yields a set of "derivative subsystems" from the overlapping regions, each having a sign corresponding to the degree of overlap in the primary subsystems. Next, a calculation is performed on the parent molecule using a low level of theory, while high-level calculations are performed for the primary and derivative subsystems. The high- and low-level energies of the subsystems are summed according to their coefficients as determined by the IEP to yield high- and low-level fragmentation energies for the system. The low-level fragmentation energy is then subtracted from the low-level energy of the real, full system, and the high-level fragmentation energy is added in to extrapolate to the high-level energy for the whole molecule. This procedure details MIM2, where the number refers to the number of levels of theory involved in the calculation (or "layers" in the ONIOM formalism). MIM, ONIOM, and AroBOROS can all be extended to an arbitrary number of levels of theory by considering intermediately sized subsystems encompassing the smaller subsystems. Herein, we focus on MIM2 calculations but do motivate a hierarchy of primary subsystems that may be used and combined to form "*n*" layer calculations.

This report will be organized as follows. The Methodology section, Section II, will detail the AroBOROS algorithm for creating subsystems for the fragmentation of cata-condensed and fully benzenoid, peri-condensed PAHs. The method used for testing the algorithm is described in Computational Details (Section III). Results and Discussion (Section IV) will begin by justifying the use of Clar structures in the interpretation of the subsystems used. We will then discuss the performance of the algorithm on two test sets of medium-to-large sized PAHs, leading up to demonstration calculations on three disparate, large PAHs with subsystems much smaller than previously considered for similar systems. We will also demonstrate the speed-up in the calculation time using the AroBOROS method compared to the full calculation and then compare our method with previous approaches. The Conclusions are listed in Section V.

II. METHODOLOGY

It is useful to begin by briefly explaining the relevant equations in the MIM method, and more details can be found in the original paper.[29] Here we are interested in MIM2, referring to two levels

https://dx.doi.org/10.1021/acs.jctc.9b00566
J. Chem. Theory Comput. 2020, 16, 2160−2171

of theory, as previous work has indicated this is the best approach when a low-level calculation on the full molecule is viable.[46] Equation 1 presents the inclusion−exclusion principle used for summing up the energies of the primary and derivative subsystems. Here, E^i represents the energy of the ith primary subsystem and $E^{i\cap j}$ corresponds to the derivative subsystem formed from the overlapping region between primary subsystems i and j. In this nomenclature, $E^{i\cap j}$ is the two-body overlap between subsystems i and j which is subtracted to avoid overcounting. The IEP is a general way of handling these overlapping terms to an arbitrary degree of overlap and becomes critically important in the discussion to follow. The subscripts l and h correspond to whether the energy evaluations are performed at the low or high level to form either the model low, *ml*, or model high, *mh*, fragmentation energies for the full system, respectively. Equation 2 is a generalization of the ONIOM extrapolation and recovers the full MIM2 energy by subtracting the model low energy, E_{ml}, from the real low-level energy of the full molecule, E_{rl}, and adding in the model high fragmentation energy, E_{mh}, as a correction.

$$E_{ml/mh} = \sum_i E^i_{l/h} - \sum_{ij} E^{i\cap j}_{l/h} + \ldots + (-1)^{n-1} \sum_n E^{i\cap j\cap k\ldots\cap n}_{l/h} \quad (1)$$

$$E^{MIM2} = E_{rl} - E_{ml} + E_{mh} \quad (2)$$

Following the MIM formalism, the embedded benzene rings of a PAH can be used as primitive fragments to assemble a set of overlapping primary subsystems. Considering just two benzene rings per subsystem, each primary subsystem may manifest as either a "bond-connected" biphenyl or a "ring-adjacent" naphthalene. Given the simplicity of the units, there may be many possible sets of biphenyl and naphthalene primary subsystems capable of filling the space of a given PAH. The goal in using such small subsystems, however, is to be able to flexibly account for aromaticity in every fragment calculation in such a way that we best represent the full PAH. With this in mind, the "correct" set of subsystems must overlap to yield derivative subsystems composed only of whole rings which guarantees aromaticity is undisturbed in every fragment calculation. The added condition of forcing the overlap between all subsystems to be only whole rings drastically reduces the number of possible tilings of primary subsystems.

To systematically solve the problem of fragmenting a large polyaromatic hydrocarbon, AroBOROS has been developed as an algorithm to determine which primary subsystems are necessary to yield a derivative subsystem set comprised entirely of whole rings. First, we must consider all possible biphenyl and naphthalene primary subsystems able to be formed in the PAH. At this point, there are too many subsystems present to maintain an overlap set composed of whole rings and we must systematically eliminate primary subsystems until such a set is found. The elimination of subsystems is determined by tracking "violations" in subsystem pairs, which constitute any overlap that is a non-whole-ring structure. Violations include not just broken rings but also structures such as styrene which are comprised of whole rings with nonwhole ring decorations. These violations must be found by determining the two- and three-body derivative subsystems for the full set of primary subsystems.

To ensure a viable fragment set, priority is designated to certain primary subsystems, and subsystems in violation with these priority subsystems are eliminated. An initialization step is taken to ensure edge subsystems take priority for the first round of eliminations. Eliminations then proceed recursively such that the least violating subsystems gain priority and their violating counterparts are eliminated. This procedure is done until there are no violations in the subsystem set. As the algorithm proceeds, it is important to consider the entire IEP as it is possible for two-body overlaps that count as violations to be canceled out in the three-body overlaps. As an example, when a naphthalene subsystem is needed for the final set it is likely that an ethene two-body derivative subsystem will occur. Typically, this non-whole-ring derivative subsystem will be canceled by an equivalent three-body derivative subsystem, and overzealousness in eliminating violations just by considering two-body overlaps can lead to an incomplete fragment set. AroBOROS tracks this type of cancellation of ethene derivative subsystems to avoid overcounting violations that would not exist if the whole IEP was applied at once. An example of this cancellation can be found in the SI.

Currently, without allowing approximations in exempting some two- and three-body violations, this algorithm applies to all fully benzenoid, peri-condensed PAHs and all cata-condensed PAHs. A peri-condensed PAH contains at least one carbon atom shared between three rings, whereas a carbon atom can at most be shared by two rings in a cata-condensed PAH. Figure 1 shows

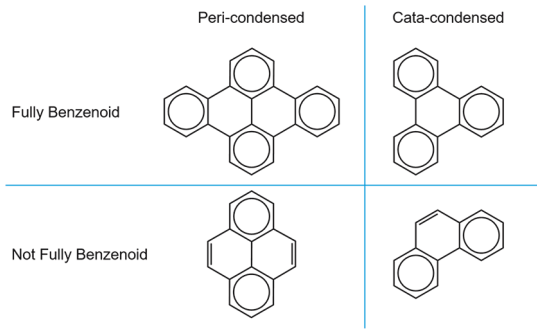


Figure 1. Examples of fully and nonfully benzenoid, peri- and cata-condensed PAHs. Clockwise from upper left: dibenzopyrene, triphenylene, phenanthrene, and pyrene.

some examples of peri- and cata-condensed PAHs; the current algorithm can address three out of four of these PAH types. The majority of peri-condensed, nonfully benzenoid PAHs cannot be spanned exactly by a fragment set containing only whole ring subsystems, and some approximations must be considered. While these approximations will be explored in the future, the majority of synthesizable and well-defined PAHs are cata-condensed and/or fully benzenoid.

The distinction between cata- and peri-condensed systems is straightforward, but identifying a fully benzenoid PAH may require some intuition. To systematically determine the fully benzenoid PAHs, we provide some guidelines: (1) the PAH must have its number of conjugated carbon atoms be divisible by 6, (2) all edges must be arm-chair, and (3) any presence of a zigzag edge feature will mean the PAH is not fully benzenoid. If the reader finds points 2 and 3 too tedious or wants to minimize human error, a program exists to characterize the resonance structures of a PAH. The ZZDecomposer program[54] generates the Zhang−Zhang polynomial[55] for a given PAH, which contains information on the maximum number of aromatic sextets in the PAH. Without going into detail on the Zhang−Zhang polynomial, the degree of the polynomial (largest

https://dx.doi.org/10.1021/acs.jctc.9b00566
J. Chem. Theory Comput. 2020, 16, 2160−2171

exponent) represents the maximum number of aromatic sextets for the PAH. For a fully benzenoid PAH, this number must be equal to the number of conjugated carbon atoms divided by 6. The Zhang−Zhang polynomial method is a general graph theory-based technique able to provide a wealth of information on the Clar structures (also called "Clar covers") of a given PAH based entirely on its connectivity, requiring no chemical intuition, and the authors highly encourage researchers in the field of large PAHs to utilize and consider this technique.

After describing the nature of systems applicable to the algorithm and how to discern them, we can now detail the full procedure. The AroBOROS algorithm yields a set of biphenyl and naphthalene primary subsystems that overlap in whole ring units. The resulting set of subsystems can be compared to the ideal Clar structures of the parent molecule; any ring with pi electrons will be involved in a primary subsystem. Empty rings in the ideal Clar structure do not need to be explicitly present in the subsystem set. The steps involved in the AroBOROS method are outlined as follows:

1. Determine all rings in the PAH from the connectivity of the atoms.
    a. Check which rings lie on the edge of the PAH.
2. Determine ring connectivity.
    a. Two possibilities are "bond-connected" or "adjacent".
        i. Bond-connected—Biphenyl
        ii. Adjacent—Naphthalene
3. Create a set of primary subsystems that take each possible biphenyl and naphthalene subsystem based on ring connectivity.
    a. Use the list of edge rings to determine which primary subsystems contain two edge rings, which will result in the designation of an edge primary subsystem.
4. Find two- and three-body overlapping derivative subsystems of the current set of primary subsystems.
    a. Compare ethene two- and three-body overlapping derivative subsystems to determine which will cancel.
5. Determine "violations" of the primary subsystems by looking through two-body derivative subsystems.
    a. A violation is defined as a non-whole-ring derivative subsystem from the two-body derivative subsystem list.
        i. This will manifest as any subsystem smaller than 6 carbon atoms, a nonring forming structure of carbon, or a nonring forming structure connected to a whole ring.
    b. Count the number of violations caused by each primary subsystem and store a list of the violating primary subsystem pairs, not counting canceled ethene derivative subsystems.
6. Initialize the system for primary subsystem elimination to reach the desired set of primary subsystems that only overlap in whole ring units.
    a. Eliminate all subsystems violating with edge naphthalene primary subsystems.
        i. Edge naphthalene subsystems will always be present in the final set of primary subsystems.
    b. Eliminate all subsystems violating with edge biphenyl primary subsystems.
7. Update violation count for the remaining subsystems.
8. Check if there are remaining violations.
9. If there are violations remaining, find the subsystems with the least number of violations and assign priority.
10. Eliminate subsystems that violate with the new priority subsystems.
11. Return to step 7 until all violations are eliminated.

To give the reader an expectation as to what subsystem sets will be used, all fully benzenoid PAHs can be represented by a collection of biphenyl subsystems, whether cata- or peri-condensed. A PAH that is not fully benzenoid will necessarily have naphthalene units in the set. An exploded view of the primary subsystems for all benchmark systems is provided in the SI.

Finally, the set of biphenyl and naphthalene units can be combined to construct larger primary subsystems by following a simple set of rules:

1. The two subsystems overlap by a whole ring.
    a. Two biphenyl subsystems must form a triphenylene subsystem.
    b. A biphenyl and naphthalene subsystem can be combined as long as the link atoms are not overlapping.

Restricting the biphenyl subsystem combinations to forming triphenylene subsystems is necessary to control the degree of overlap of the subsystems. This establishes a hierarchy in the subsystem sizes, with the smaller, two ring subsystems belonging to "Rung 1" and their valid combinations forming "Rung 2". Figure 2 shows the possible subsystems for Rung 1 and Rung 2.

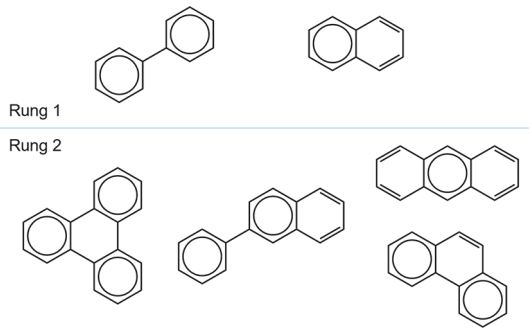


Figure 2. Rung 1 subsystems: biphenyl and naphthalene. Rung 2 subsystems: triphenylene, phenylnaphthalene, anthracene, and phenanthrene.

We hope to motivate a hierarchy of additional fragments up through Rung 6 by continuously adding biphenyl or naphthalene fragments and will pursue this in future work.

III. COMPUTATIONAL DETAILS

Geometries of molecules with 84 or more carbon atoms were optimized in the MOPAC semiempirical program package[56] with the dispersion-corrected PM6-D3H4 method.[57] All other electronic structure calculations in this work were performed using the development version of the Gaussian program suite, Gaussian.[58] Geometries of molecules with less than 84 carbon atoms were optimized at the M06-2X/6-31+G(d,p) level of theory.[59] All reported MIM2 calculations are at the [M06-2X/6-311+G(2df,p):M06-2X/6-31+G] level of theory where the colon demarcates the high and low levels of theory as illustrated in eqs 1 and 2. We chose M06-2X for all energy evaluations due to its excellent performance in predicting the thermochemistry

https://dx.doi.org/10.1021/acs.jctc.9b00566
J. Chem. Theory Comput. 2020, 16, 2160−2171

of pi-conjugated systems and its ubiquity in this area of research.[60] All DFT calculations used the ultrafine grid in Gaussian. Errors for each molecule are reported as the absolute difference between the extrapolated MIM2 and full high-level energies.

AroBOROS was implemented as an external Perl script to create primary subsystems for the MIM Perl suite, which interfaces with Gaussian. Due to the large degree of overlap in the initial set of primary subsystems, the determination of derivative subsystems was performed using a bit-manipulation algorithm obtained from the Iyengar group at Indiana University. This subsystem-determination algorithm was originally devised to enable high speed determination of fragments at each step of fragment-based *ab initio* molecular dynamics calculations.[61,62] Reference 62 contains pseudocode describing this algorithm, and the derivative subsystem portion of this code was interfaced with the AroBOROS script to decrease the scaling of this step.

IV. RESULTS AND DISCUSSION

A. Triphenylene and Tetracene. Testing potential fragmentation schemes for fully benzenoid, cata-condensed PAHs, we notice the primary subsystem set yielding only whole ring overlap can consist of either entirely biphenyl or naphthalene units. *A priori*, considering Clar stuctures suggests the biphenyl fragment set will better represent the system, and this can be demonstrated by analyzing the potential AroBOROS fragmentation schemes for the PAH triphenylene. Figure 3

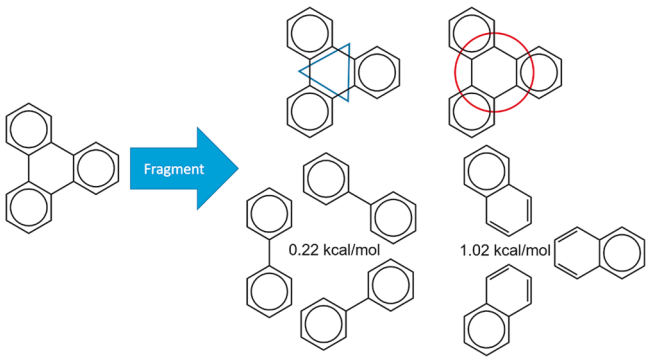


Figure 3. Biphenyl and naphthalene primary subsystem sets for fragmenting triphenylene. The blue triangle and red circle indicate which bonds are cut to obtain the two fragment schemes. The biphenyl subsystem set produces an error 0.80 kcal/mol lower than the naphthalene set.

presents the Clar structures for triphenylene and its representative fragment sets. The biphenyl set clearly reproduces the Clar structure of triphenylene while the naphthalene set assigns some sextet character to the central empty ring. Additionally, cutting the bonds crossed by the blue triangle in Figure 3 forms the biphenyl set while one must cut the bonds under the red circle to form the naphthalene set. The naphthalene fragment set requires six fully aromatic bonds to be cut while no sextets are disturbed to obtain the biphenyl set. Fittingly, the biphenyl set matches the high-level calculation with 0.8 kcal/mol less error than the naphthalene set. The dominance of the biphenyl set extends to all cata-condensed, fully benzenoid PAHs; thus, this set is always selected over the corresponding naphthalene set. Fully benzenoid, peri-condensed PAHs can only be represented by a set of biphenyls, and attempting to construct a fragment set of naphthalenes will always result in non-whole-ring overlaps.

Tetracene possesses the same number of carbon atoms as triphenylene but requires a different set of subsystems for its fragmentation scheme. Figure 4 illustrates the migration of the

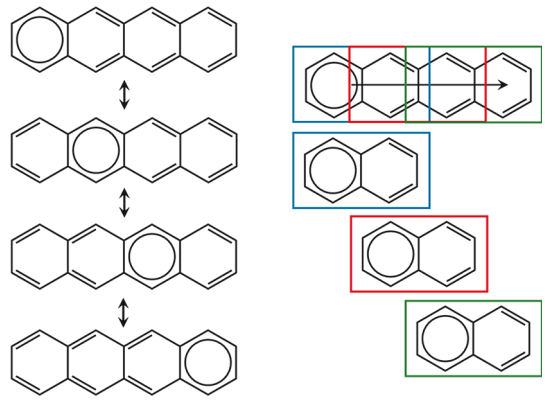


Figure 4. Resonance structures of tetracene (left). Simplified resonance notation and color coordinated naphthalene primary subsystems (right). Drawing an arrow from the aromatic sextet is shorthand for the migrating resonance structures on the left.

delocalized aromatic sextet in the Clar representation of tetracene. Every ring in tetracene has some contribution from a full sextet, and the appropriate fragment set must take this into account. Figure 4 shows the three overlapping naphthalene units making up the only possible AroBOROS fragment set for tetracene. Inevitably, bonds with some contribution from an aromatic sextet will be cut, but this does not lead to significant error (0.14 kcal/mol) when compared to the high-level calculation on the full system. The low error in tetracene further cements the importance in considering the Clar structure of individual fragments when assessing the performance of fragmentation methods on PAHs.

B. Cata-Condensed PAHs. Figure 5 shows the cata-condensed set of molecules used to test the fragmentation method. CC in the benchmark set labels corresponds to "cata-condensed", followed by the number of carbon atoms and FB if the species is a fully benzenoid species. Four of the benchmark

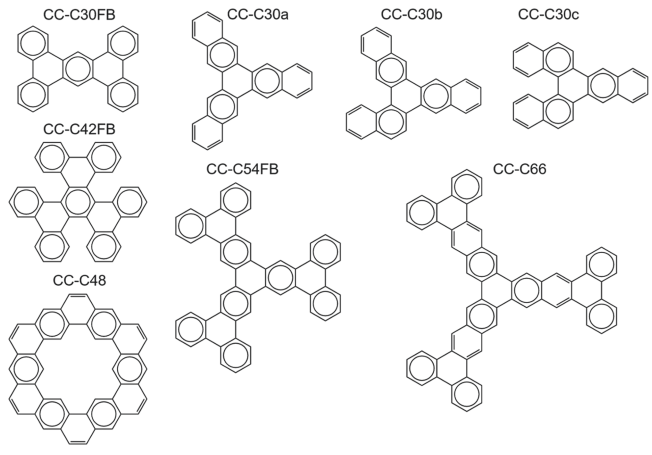


Figure 5. Test set of cata-condensed PAHs, both fully benzenoid and not. The naming convention is "CC" for cata-condensed, "C#" for the number of carbon atoms, and "FB" if the system is fully benzenoid. The dominant Clar structure is drawn for each system.

https://dx.doi.org/10.1021/acs.jctc.9b00566
J. Chem. Theory Comput. 2020, 16, 2160−2171

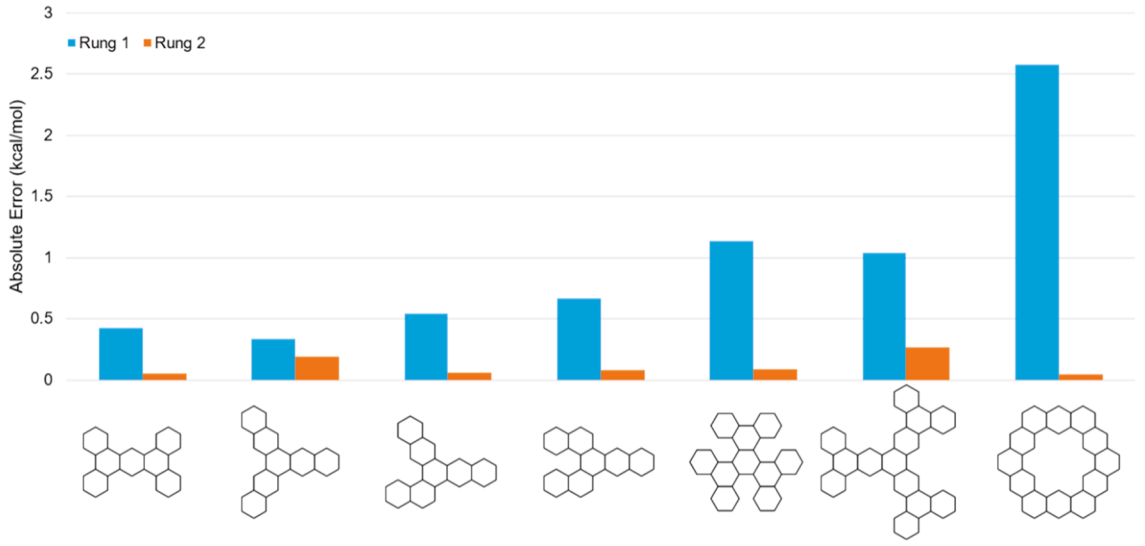


Figure 6. Absolute errors of cata-condensed PAHs. The blue bars refer to errors for the Rung 1 subsystems while the orange bars refer to Rung 2 subsystems.

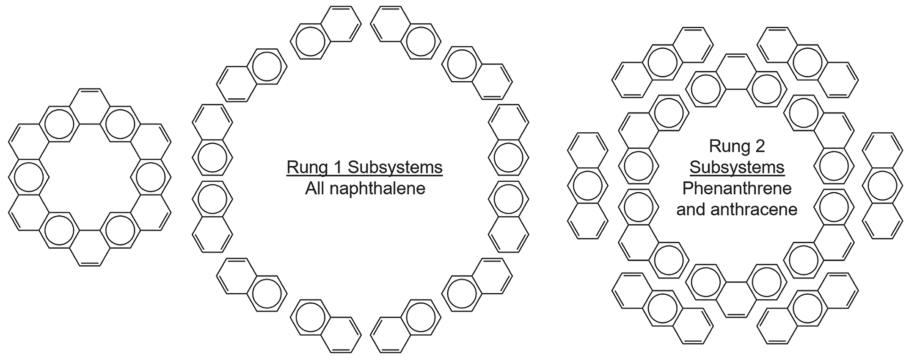


Figure 7. Exploded views of Rung 1 and 2 subsystems for the fragmentation of CC-C48.

set PAHs contain 30 carbons, three of these being not fully benzenoid and differing only by the connectivity of rings on a central triphenylene unit. CC-C48 (kekulene) was the first example of the cycloarene class of PAHs, which has been important for studying the Clar convention in aromaticity[63,64] and served as a potential model for holes in carbon materials.[65] The CC-C66 molecule ([16]cloverphene(5.5.5)) is the base structure of one of the largest cata-condensed PAHs ever synthesized.[66]

Figure 6 shows the errors for each PAH in the cata-condensed test set. Rung 1 produces errors less than 1.5 kcal/mol for every system except CC-C48, kekulene. Moving up to the slightly larger subsystems of Rung 2, errors drop to below 0.5 kcal/mol for the entire set. The SI contains diagrams outlining the fragmentation of each test system, and one can easily verify the better representation of the Clar structures presented in Figure 5 at Rung 2 as opposed to Rung 1. The most dramatic example of improvement from Rung 1 to Rung 2 is kekulene, which at first seems strange given this is neither the largest nor the most strained system in the set. Figure 7 shows the subsystem sets used in the Rung 1 and Rung 2 calculations for CC-C48. The fully naphthalene fragment set generated by Rung 1 does a poor job of approximating the full kekulene structure due to its lack of rings containing only one pair of pi electrons. Rung 2, on the other hand, contains phenanthrene units representing a much better local approximation of kekulene's Clar structure, leading to a much more favorable error while each 14 carbon atom fragment is only 29% the size of the full 48 carbon atom system.

Several systems in the cata-condensed set examine how AroBOROS performs when strain is introduced into the system, and Figure 8 presents 3D models of these strained systems. The

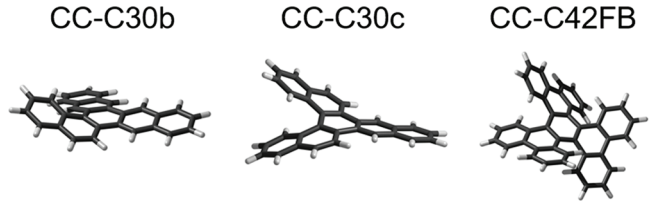


Figure 8. 3D models of CC-C30b, CC-C30c, and CC-C42FB, demonstrating the strained geometries.

CC-C30a-c series provides some useful insight on how systematically increasing the strain of the system affects the performance of the method. The error grows steadily with increasing steric strain from forcing hydrogen link atoms near each other, which can be observed by comparing the Rung 1 absolute errors for CC-C30a-c. Even for a highly strained system, CC-C42FB, the Rung 1 absolute error (1.13 kcal/mol) is excellent, especially when taking into account a single biphenyl fragment, 12 carbon atoms, only represents 29% of the full system.

https://dx.doi.org/10.1021/acs.jctc.9b00566
J. Chem. Theory Comput. 2020, 16, 2160−2171

C. Peri-Condensed Fully Benzenoid PAHs. Figure 9 includes a test set of 5 fully benzenoid, peri-condensed PAHs.

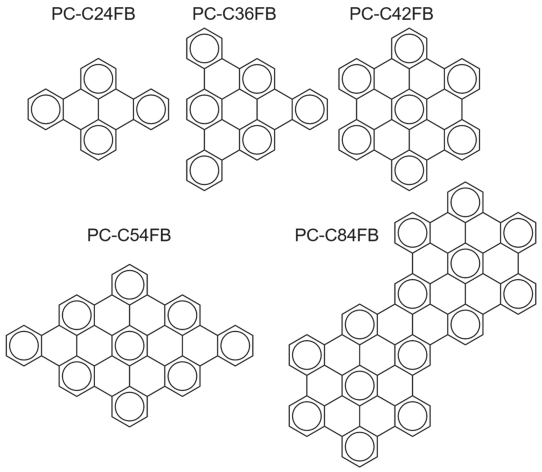


Figure 9. Fully benzenoid, peri-condensed benchmark PAHs. The naming convention is "PC" for peri-condensed, "C#" for the number of carbon atoms, and "FB" if the system is fully benzenoid. The dominant Clar structure is drawn for each system.

Due to their stability, the PAHs most often synthesized when targeting graphene quantum dots are peri-condensed and fully benzenoid. PC-C24FB is the smallest of these species while PC-C84FB is the unit cell structure of chevron graphene nanoribbons reported to span over 500 nm in length.[67] As mentioned in the discussion of triphenylene in part A of this section, the Rung 1 subsystem set for each molecule in Figure 9 contains only biphenyl while Rung 2 forms only triphenylene. Errors for each system versus the high level can be seen in Figure 10, and one can notice a near linear growth in the error of Rung 1 as the system size increases. Applying the Rung 2 subsystems (all triphenylene units) reduces the error greatly, with no error above 1.5 kcal/mol. The chevron unit cell, PC-C84FB, represents the largest Rung 2 error but still falls within the 2 kcal/mol threshold of chemical accuracy while each triphenylene subsystem, 18 carbon atoms, only makes up 21% of the full system. The Clar structures of the fully benzenoid species clearly show the molecules can be viewed as a series of embedded biphenyl or triphenylene units.

D. Comparing Cata- and Peri-Condensed Performance. There are two useful comparisons to make between the cata- and peri-condensed test systems, namely, between the systems that have 42 and 54 carbon atoms in each set. We can make direct comparisons in the error for systems with stoichiometrically equivalent numbers of carbons. Table 1 shows the error for the AroBOROS method in each system, which is only appreciable in the Rung 1 case. In each case, the cata-condensed PAH has a lower error than the corresponding peri-condensed PAH. Peri-condensed PAHs will have more aromatic bonds than the corresponding cata-condensed system with an equivalent carbon stoichiometry, and each aromatic bond cleaved in the fragmentation process contributes to the overall error. Using Rung 2 subsystems largely eliminates the differences, and both cata- and peri-condensed systems produce errors below 0.5 kcal/mol.

Table 1. Absolute Errors of 42 and 58 Carbon Atom Systems[a]

	Rung 1		Rung 2	
# carbon atoms	cata	peri	cata	peri
42	1.13	1.95	0.09	0.16
54	1.04	2.69	0.27	0.31

[a]Errors in kcal/mol relative to the full high level calculation. Represented systems are CC-C42FB, CC-C54FB, PC-C42FB, and PC-C54FB.

The impact of breaking more aromatic bonds is evident when comparing the errors as they change while growing cata- and peri-condensed PAHs. Figure 11 plots the Rung 1 absolute errors for both the cata- and the peri-condensed systems. In general, the peri-condensed systems have a higher rate of error growth with increasing system size when compared to the cata-

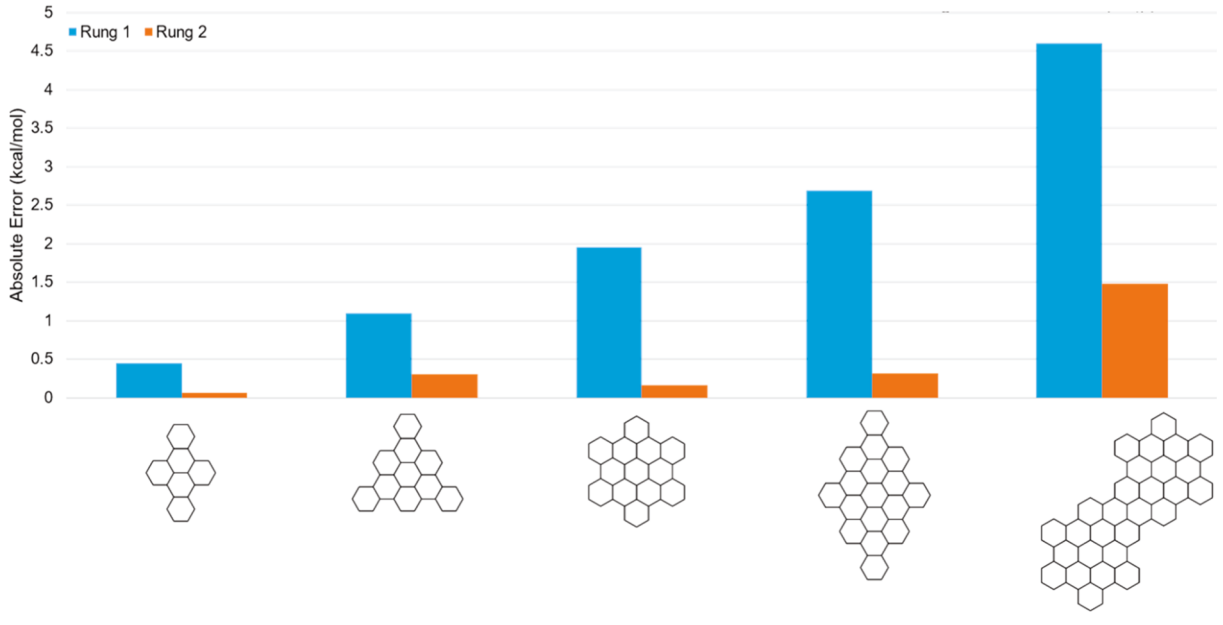


Figure 10. Absolute errors of peri-condensed PAHs. The blue bars refer to errors for the Rung 1 subsystems and the orange bars refer to Rung 2.

https://dx.doi.org/10.1021/acs.jctc.9b00566
J. Chem. Theory Comput. 2020, 16, 2160−2171

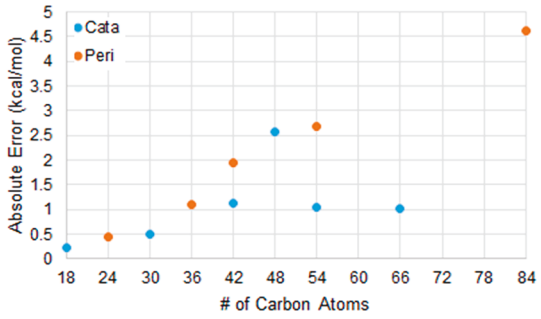


Figure 11. Absolute error with Rung 1 subsystems versus the number of carbon atoms in each system. The blue dots correspond to the cata-condensed systems, and the orange dots correspond to the peri-condensed systems.

condensed systems. This is likely due to the larger number of aromatic bonds broken in the fragmentation of the peri-condensed systems.

E. Large Molecules. We apply our method to an appreciably large graphene quantum dot that has been synthesized and characterized in the lab (PC-C132FB),[68] a truncated armchair carbon nanotube with 108 carbon atoms (SWNT(6,6)-C108FB), and one of the largest cata-condensed PAHs synthesized (CC-C102).[66] The nomenclature for the nanotube derives from the standard notation for single wall nanotubes (SWNT) where (6,6) refers to the dimensions $m = n = 6$, indicating the width of the nanotube and its armchair structure. This specific, truncated nanotube also appeared in the recent MFCC paper on fragmenting conjugated systems, where the subsystems in that work represented 66.7% of the total structure.[42] The structures of these three large PAHs are shown in Figure 12 and the absolute errors summarized in Table 2.

The absolute errors for PC-C132FB and SWNT(6,6)-C108FB with Rung 1 subsystems are quite large, which is to be expected due to their size. The error of the fragmentation scheme is an extensive property of the system and should grow with system size in a strongly delocalized system. The improvement from using the Rung 2 subsystems is significant, exhibiting a reduction of error by 6 to 11 kcal/mol. The nanotube fragment shows the most improvement, indicating that increasing the size of the subsystems does well to describe the strained system. CC-C102 does reasonably well even with the Rung 1 subsystem and has nearly perfect performance with the Rung 2 subsystems. The excellent representation of the large, cata-condensed PAH arises due to less aromatic bonds being broken in the fragment set; this effect was observed previously with the comparison of cata- and peri-condensed

Table 2. AroBOROS Absolute Errors of Three Large PAHs[a]

system	Rung 1	Rung 2
PC-C132FB	9.43	3.25
SWNT(6,6)-C108FB	14.22	3.26
CC-C102	3.06	0.22

[a]Absolute errors relative to full high level calculation (kcal/mol).

systems. The percent of carbon atoms in the largest fragment for each system at Rung 2 is only 13.6% for the fully benzenoid C132, 16.7% for the nanotube C108, and 17.6% for cloverphene, C102. More specifically, the largest subsystem we use (at the high level of theory) for any molecule in this report is triphenylene, constituting 18 carbon atoms, and is independent of the size of the parent system, enabling applications to larger PAHs. Indeed, the performance illustrated in Table 2 indicates even small subsystems can be used to obtain accurate results for appreciable large PAHs.

F. Representative Timings. In order to assess the speedup when using Rung 2 AroBOROS subsystems compared to the full high-level calculations, 6 representative molecules (3 cata-condensed and 3 peri-condensed) were chosen and rerun serially on a single machine in a controlled timing environment with identical computational resources available for each calculation. The results of these timings can be seen in Table 3 as well as the number of Rung 2 subsystems (# R2) and the

Table 3. Timings and Relative Speed-up of AroBOROS with and without Symmetry-Redundant Subsystems[a]

System	High (u)	AroBOROS (u)	RL (u)	# R2	Speed-up (×)
PC-C42FB	15.58	8.82 [2.31]	1.00	6 [1]	**1.77** [6.74]
PC-C84FB	82.02	24.55 [13.27]	1.48	14 [7]	**3.21** [6.18]
PC-C132FB	292.25	55.70 [38.52]	20.26	26 [13]	**5.25** [7.59]
CC-C42FB	10.83	5.08 [2.61]	1.31	3 [1]	**2.13** [4.15]
CC-C54FB	24.27	5.77 [3.72]	1.78	4 [2]	**4.21** [6.52]
CC-C66	31.63	18.28 [6.64]	2.43	16 [4]	**1.73** [4.76]

[a]Timings run serially on the same machine, reported in units (u) of the timing of the full, real low calculation of PC-C42FB. Bolded and blue values in brackets correspond to timings with symmetry taken into account and identical subsystems only run once. High corresponds to the full, high level calculation, and AroBOROS refers to the overall serial timing of the fragmentation method (all model low and model high calculations as well as the real low calculations summed together), RL to the full low-level calculation on the entire molecule, # R2 to the number of Rung 2 subsystems, and Speed-up to High/AroBOROS: the multiplicative speed-up when using the AroBOROS method. Levels of theory as reported in Methodology.

speed-up (defined as the full, high-level calculation divided by the AroBOROS timing with all components computed serially on the same machine). The time of the shortest "real low" (RL)

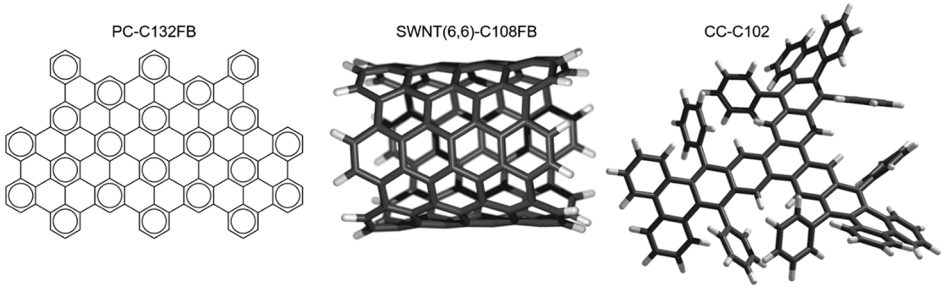


Figure 12. Structures of the large systems to test Rungs 1 and 2 of the AroBOROS method.

https://dx.doi.org/10.1021/acs.jctc.9b00566
J. Chem. Theory Comput. 2020, 16, 2160−2171

calculation (that of PC-C42FB) was chosen as a normalized unit (u) applied to all reported timings to make results transferable to and testable on other machines. Due to the symmetric nature of the systems investigated throughout this study, Table 3 also presents timings and quantities derived from calculations where subsystems that would be redundant by symmetry are removed (the blue and bolded quantities in brackets). The bracketed values in the R2 column of Table 3 presents the number of symmetry unique subsystems. These values place the fragmentation-based calculations on more even footing with the high-level calculations due to the use of symmetry in the full, reference calculations.

Starting with the peri-condensed systems reported in the first 3 rows of Table 3, it is clear that speed-up increases with system size, leading to an overall speed-up of 5.25× (without using any symmetry) for PC-C132FB, the largest molecule included in this study. Examining the real low (RL) timings for the peri-condensed systems, it is also noticeable that this calculation becomes the computational bottleneck as system size increases, motivating the 3-layer work we have recently reported as a solution to this problem.[48] This being said, however, reporting the full AroBOROS timing as well as the real low timing makes it clear that running these calculations in parallel will net substantial speed-ups, even if only the fragment and real low calculations are run separately. Taking symmetry into account leads to large speedups even in the serial case, with PC-C132FB netting a speed-up of 7.59× simply by taking the mirror plane of the molecule into account.

Analysis becomes slightly more complicated when considering the cata-condensed subsystems. Similar behavior is observed for the fully benzenoid cata-condensed subsystems as seen in the fully benzenoid peri-condensed subsystems, with speed-up increasing rapidly with system size and the total number of Rung 2 subsystems (# R2) kept quite low. However, in the nonfully benzenoid case, CC-C66, more subsystems are required due to the connectivity of nonfully benzenoid subsystems and speed-up is only modest (1.73× without using any symmetry). This is an artifact of the total number of subsystems required to treat these branching molecules; however, this can be greatly assuaged by running subsystems in parallel as opposed to serially. However, we note that significant speed-up can still be achieved in the serial sense for systems such as CC-C66 by using symmetry to determine which subsystems are unnecessary (i.e., the speed-up of 4.76× shown in the table where the only required Rung 2 subsystems are an inner triphenylene subsystem, an outer triphenylene, an inner phenylnaphthalene, and an outer phenylnaphthalene). These improvements, shown in Table 3, are easily obtained from AroBOROS subsystems via postprocessing.

G. Comparison with Previous Methods. As alluded to in the Introduction, fragmentation methods have been applied to graphitic and conjugated systems previously. Here we will compare the AroBOROS methodology with those used previously, specifically with an eye on how fragments are chosen and their sizes, as well as the final speed-ups and goals of each method. This work focuses explicitly on polycyclic aromatic hydrocarbons, so we will focus the discussion on the performance of each method to this class of molecules. The pioneering work in applying fragmentation methods to polycyclic aromatic hydrocarbons and to conjugated molecules in general was reported by Yeole and Gadre with the MTA fragmentation approach.[40] This work employed a single level of theory (as do all methods reported in this section aside from MIM), necessitating relatively large fragments to produce accurate single-point energies for PAH systems ranging in size up to a 56-carbon graphene flake (which we will refer to as C56). Two fragmentation schemes were reported for the calculation of single-point energies, **S1** and **S2**, with **S1** consisting of smaller fragments than **S2**.

In the case of C56, the less-accurate scheme **S1** takes the form of two C40 fragments (what we would call primary subsystems, in this work) overlapping to form a third fragment (derivative subsystem) with 20 carbons. The nature of the **S2** fragmentation for C56 is not explicitly detailed but described with respect to the Rg (R-goodness) parameter, which is defined as the minimum value over all atoms in a given fragment of the maximal radius of the sphere centered on any atom within a fragment that can be drawn and include only other atoms within that fragment. The Rg parameter of **S1** for C56 (consisting of two C40 primary subsystems) was reported as 4.9 Å while the Rg of **S2** was reported as 6.1 Å. Both B3LYP/6-31+G and M06/6-31+G errors were reported and were very similar throughout the study, with C56 B3LYP errors, relative to the high-level, reported as 2.14 and −0.41 kcal/mol for schemes **S1** and **S2**, respectively. In terms of timings, the high-level calculation with B3LYP was reported to take 470 min, while the **S1** MTA treatment was reported to take 560 min: 1.19 times longer than the full calculation. The timing of the **S2** scheme for C56 was not reported in this work.

Due to the necessity of large fragments when considering PAHs, Yeole and Gadre turned to geometry optimization, noting that gradient estimations for small fragmentation schemes can be quite good even when energies are in error. With this in mind, fragmentation schemes **S3** and **S4** were developed to produce smaller fragments with higher speed-ups but worse single-point energies. These schemes were employed for C56 and a 96-carbon graphene flake, C96, for which single-point energies were not reported, and were found to reproduce geometry optimization at the high-level to a high degree of accuracy.[40]

Huang, Bohorquez, Matta, and Massa extended the KEM many-body fragmentation approach to graphitic systems shortly after the initial MTA paper.[41] The major innovation of this work was the realization that accurate fragmentation energies could be obtained by "fissioning" aromatic bonds. This fission technique of fragment selection produced fragments that overlap by whole rings and can be considered an antecedent of the work presented here. The method was tested on the single-point energy of a 78-carbon graphene flake, with the main fragmentation consisting of two 54-carbon "double kernels" (analogous to primary subsystems, here). The method was tested against HF/3-21G and MP2/3-21G with errors of −1.2 and 0.94 kcal/mol, respectively. Timings were not reported in this work.[41]

Finally, the most recent application of fragmentation techniques to PAHs was reported by Zhang with the MFCC method.[42] This work reported interaction energies of molecules (water, atomic cations, and CO) inside of and interacting with C60, carbon nanotubes, and a 270-carbon graphene sheet. Fragments in this study were arrived at by cutting bonds and "capping" with the appropriate geometry of carbons to fulfill aromaticity, resulting in fragments similar to the KEM study: slices of materials overlapping at a layer of whole rings. C60 was fragmented into two halves, CNTs were fragmented into cylindrical segments preserving the full curved nature of the nanotube, and the graphene sheet was fragmented into four 90-carbon flakes overlapping at a single layer of whole rings.

https://dx.doi.org/10.1021/acs.jctc.9b00566
J. Chem. Theory Comput. 2020, 16, 2160−2171

Focusing on graphene, timings were reported for the average fragment vs the time required for the full system at the B3LYP/6-31G* level of theory, with the full system requiring 2741 min and the average fragment requiring 341, for a speed-up of 8 times. If all fragments (minus the overlapping region) are considered, this reduces to a speed-up of 2 times, but focusing on average fragment times is not dissimilar to running all fragments in parallel on separate machines. It should also be noted that schemes breaking the 270-carbon sheet into 6 fragments were also discussed but timings not presented, with the scheme yielding the lowest number of carbons per fragment containing two 72-carbon fragments and four 54-carbon fragments. Single-point energies were not reported, however interaction energy curves showed excellent agreement throughout.[42]

Considering the current work, we endeavored to develop a systematic algorithm for the fragmentation of large PAHs that tessellates a controlled set of overlapping subsystems to prevent fragment sizes from creeping larger as the system does. The largest fragment employed in this work is triphenylene, with 18-carbon atoms. Summarizing the highlights of previous work: a 56-carbon graphene flake was represented with 40-carbon fragments each constituting 71% the size of the full system, a 78-carbon flake was represented by fragments of 56 carbons each constituting 72% of the full system, and a 270-carbon sheet was represented by fragments of 90 carbons (each constituting 33% of the full system). We present two-layer fragmentation of a 54-carbon flake represented by 18-carbon fragments (each making up 33% of the total system), an 84-carbon flake with each 18-carbon fragment making up 21% of the full system, and a 132-carbon flake with each 18-carbon fragment making up 14% of the full system, among a wide test set of PAH molecules. Capping the fragment size, combined with multiple levels of theory, allows us to achieve net speed-ups as well as maximum opportunity for parallelization and symmetry consideration (as shown in the previous section) to dramatically increase net speed-up over the full calculation. In particular, we show single-point energies *can* be calculated with small fragments for large speed-ups with low errors as long as a systematic approach is taken with multiple levels of theory. Future studies on these systems will focus on geometry optimization, interaction energies, and property evaluations (including IR,[50] Raman,[53] NMR,[69] VCD,[49] and ROA[51]).

V. CONCLUSIONS

We have introduced AroBOROS, a systematic method of fragmenting PAHs with biphenyl and naphthalene subsystems in such a way that all derivative subsystems are complete rings. Additionally, we have developed a method for combining the "Rung 1" biphenyl and naphthalene fragments to systematically form larger subsystems we refer to as Rung 2. Our method begins to address the fragmentation of these systems in a systematic manner and can provide state of the art results on cata-condensed and fully benzenoid peri-condensed PAHs. It has been demonstrated that the key to achieving high accuracy with small subsystems, thus allowing maximum speed-up compared to the full calculation, is to retain characteristics of the dominant Clar structures of a given PAH.

In general, the Rung 1 AroBOROS subsystems perform well for PAHs up to 40 carbon atoms in size, while Rung 2 is recommended for larger systems. Higher rungs are currently being pursued to further improve the performance for very large PAHs. Given the stability of the fully benzenoid structures, this method applies to the majority of synthetically viable graphene quantum dots. To treat those theoretical polyhexes representing less stable or even unstable structures, we plan to further generalize our algorithm to include nonwhole ring overlapping systems that cancel each other to a significant level.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.9b00566.

History of the AroBOROS name, example of ethene cancellation, and fragment diagrams of all molecules tested (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Krishnan Raghavachari − *Department of Chemistry, Indiana University, Indiana 47405, United States;* orcid.org/0000-0003-3275-1426; Email: kraghava@indiana.edu

Authors

Benjamin W. Noffke − *Department of Chemistry, Indiana University, Indiana 47405, United States*

Daniel Beckett − *Department of Chemistry, Indiana University, Indiana 47405, United States;* orcid.org/0000-0003-4833-2269

Liang-shi Li − *Department of Chemistry, Indiana University, Indiana 47405, United States;* orcid.org/0000-0001-7566-8492

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.9b00566

Author Contributions

[†]B.W.N. and D.B. contributed equally to this work.

Funding

This work was supported in part by funds from the National Science Foundation Grants CHE-1665427 (K.R. and D.B.) and CBET-1610712 (L.L and B.W.N.) as well as the IU President's Diversity Dissertation Fellowship (B.W.N.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Cody Haycraft, Dr. Junjie Li, and Prof. Srinivasan S. Iyengar for allowing the use of their code to determine the overlapping fragments. The authors also thank the Big Red II supercomputing center for computer time.

■ REFERENCES

(1) Barnes, E. C.; Petersson, G. A.; Montgomery, J. A.; Frisch, M. J.; Martin, J. M. L. Unrestricted coupled cluster and Brueckner doubles variations of W1 theory. *J. Chem. Theory Comput.* **2009**, *5*, 2687−2693.

(2) Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kallay, M.; Gauss, J. W3 theory: robust computational thermochemistry in the kJ/mol accuracy range. *J. Chem. Phys.* **2004**, *120*, 4129−4141.

(3) Bomble, Y. J.; Vazquez, J.; Kallay, M.; Michauk, C.; Szalay, P. G.; Csaszar, A. G.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated ab initio thermochemistry. II. Minor improvements to the protocol and a vital simplification. *J. Chem. Phys.* **2006**, *125*, No. 064108.

(4) DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The correlation consistent composite approach (ccCA): an alternative to the Gaussian-n methods. *J. Chem. Phys.* **2006**, *124*, 114104.

(5) Feller, D.; Dixon, D. A. Predicting the heats of formation of model hydrocarbons up to benzene. *J. Phys. Chem. A* **2000**, *104*, 3048−3056.

(6) Harding, M. E.; Vazquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. High-accuracy extrapolated ab initio thermochemistry. III. Additional improvements and overview. *J. Chem. Phys.* **2008**, *128*, 114111.
(7) Karton, A.; Martin, J. M. L. Performance of W4 theory for spectroscopic constants and electrical properties of small molecules. *J. Chem. Phys.* **2010**, *133*, 144102.
(8) Karton, A.; Parthiban, S.; Martin, J. M. L. Post-CCSD(T) ab initio thermochemistry of halogen oxides and related hydrides XOX, XOOX, HOX, XOn, and HXOn, (X = F, Cl), and evaluation of DFT methods for these systems. *J. Phys. Chem. A* **2009**, *113*, 4802−4816.
(9) Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. W4 theory for computational thermochemistry: in pursuit of confident sub-kJ/mol predictions. *J. Chem. Phys.* **2006**, *125*, 144108.
(10) Karton, A.; Taylor, P. R.; Martin, J. M. L. Basis set convergence of post-CCSD contributions to molecular atomization energies. *J. Chem. Phys.* **2007**, *127*, No. 064104.
(11) Kowalski, K.; Piecuch, P. The method of moments of coupled-cluster equations and the renormalized CCSD T, CCSD(T), CCSD(TQ), and CCSDT(Q) approaches. *J. Chem. Phys.* **2000**, *113*, 18−35.
(12) Martin, J. M. L.; de Oliveira, G. Towards standard methods for benchmark quality ab initio thermochemistry - W1 and W2 theory. *J. Chem. Phys.* **1999**, *111*, 1843−1856.
(13) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. A complete basis set model chemistry. 5. Extensions to six or more heavy atoms. *J. Chem. Phys.* **1996**, *104*, 2598−2619.
(14) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Allaham, M. A.; Shirley, W. A.; Mantzaris, J. A complete basis set model chemistry. 1. The total energies of closed-shell atoms and hydrides of the 1st-row elements. *J. Chem. Phys.* **1988**, *89*, 2193−2218.
(15) Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musial, M. Efficient computer implementation of the renormalized coupled-cluster methods: The R-CCSD T, R-CCSD(T), CR-CCSD T, and CR-CCSD(T) approaches. *Comput. Phys. Commun.* **2002**, *149*, 71−96.
(16) Tajti, A.; Szalay, P. G.; Csaszar, A. G.; Kallay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vazquez, J.; Stanton, J. F. HEAT: High accuracy extrapolated ab initio thermochemistry. *J. Chem. Phys.* **2004**, *121*, 11599−11613.
(17) 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.
(18) Jung, Y.; Head-Gordon, M. A fast correlated electronic structure method for computing interaction energies of large van der Waals complexes applied to the fullerene-porphyrin dimer. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2831−2840.
(19) Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.* **2007**, *28*, 2618−2640.
(20) Bykov, D.; Kjaergaard, T. The GPU-enabled divide-expand-consolidate RI-MP2 method (DEC-RI-MP2). *J. Comput. Chem.* **2017**, *38*, 228−237.
(21) 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.
(22) Raghavachari, K.; Saha, A. Accurate composite and fragment-based quantum chemical models for large molecules. *Chem. Rev.* **2015**, *115*, 5643.
(23) 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.
(24) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. Molecular tailoring approach for simulation of electrostatic properties. *J. Phys. Chem.* **1994**, *98*, 9165−9169.
(25) 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.
(26) Furtado, J. P.; Rahalkar, A. P.; Shanker, S.; Bandyopadhyay, P.; Gadre, S. R. Facilitating minima search for large water clusters at the MP2 level via molecular tailoring. *J. Phys. Chem. Lett.* **2012**, *3*, 2253−2258.
(27) Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. Molecular tailoring approach in conjunction with MP2 and Ri-MP2 codes: A comparison with fragment molecular orbital method. *J. Comput. Chem.* **2010**, *31*, 2405−2418.
(28) Li, W.; Li, S.; Jiang, Y. Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules. *J. Phys. Chem. A* **2007**, *111*, 2193−2199.
(29) 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.
(30) 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.
(31) 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.
(32) 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.
(33) 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.
(34) Huang, L. L.; Massa, L.; Karle, J. Kernel energy method illustrated with peptides. *Int. J. Quantum Chem.* **2005**, *103*, 808−817.
(35) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body expansion for large systems, with applications to water clusters. *J. Chem. Theory Comput.* **2007**, *3*, 46−53.
(36) Wen, S.; Nanda, K.; Huang, Y.; Beran, G. J. O. Practical quantum mechanics-based fragment methods for predicting molecular crystal properties. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7578−7590.
(37) Řezáč, J.; Salahub, D. R. Multilevel Fragment-Based Approach (MFBA): a novel hybrid computational method for the study of large molecules. *J. Chem. Theory Comput.* **2010**, *6*, 91−99.
(38) 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*, No. 064113.
(39) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.: THEOCHEM* **1999**, *461−462*, 1−21.
(40) Yeole, S. D.; Gadre, S. R. On the applicability of fragmentation methods to conjugated π systems within density functional framework. *J. Chem. Phys.* **2010**, *132*, No. 094102.
(41) Huang, L.; Bohorquez, H. J.; Matta, C. F.; Massa, L. The kernel energy method: application to graphene and extended aromatics. *Int. J. Quantum Chem.* **2011**, *111*, 4150−4157.
(42) Zhang, D. Quantum mechanical calculation of nanomaterial-ligand interaction energies by molecular fractionation with conjugated caps method. *Sci. Rep.* **2017**, *7*, 44645.
(43) Clar, E. *The aromatic sextet*; J. Wiley: 1972.
(44) Van der Linde, J.; Havinga, E. The addition of bromine to phenanthrene in methanol as a solvent. *Recueil des Travaux Chimiques des Pays-Bas* **1965**, *84*, 1047−1058.
(45) Gross, L.; Mohn, F.; Moll, N.; Schuler, B.; Criado, A.; Guitián, E.; Peña, D.; Gourdon, A.; Meyer, G. Bond-order discrimination by atomic force microscopy. *Science* **2012**, *337*, 1326−1329.
(46) Saha, A.; Raghavachari, K. Analysis of different fragmentation strategies on a variety of large peptides: implementation of a low level of theory in fragment-based methods can be a crucial factor. *J. Chem. Theory Comput.* **2015**, *11*, 2012−2023.
(47) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions.

https://dx.doi.org/10.1021/acs.jctc.9b00566
J. Chem. Theory Comput. 2020, 16, 2160−2171

A test for Diels−Alder reactions and Pt(P(t-Bu)3)2 + H2 oxidative addition. *J. Phys. Chem.* **1996**, *100*, 19357−19363.
(48) Thapa, B.; Beckett, D.; Jovan Jose, K. V.; Raghavachari, K. Assessment of fragmentation strategies for large proteins using the multilayer molecules-in-molecules approach. *J. Chem. Theory Comput.* **2018**, *14*, 1383−1394.
(49) Jose, K. V. J.; Beckett, D.; Raghavachari, K. Vibrational circular dichroism spectra for large molecules through molecules-in-molecules fragment-based approach. *J. Chem. Theory Comput.* **2015**, *11*, 4238−4247.
(50) Jose, K. V. J.; Raghavachari, K. Evaluation of energy gradients and infrared vibrational spectra through molecules-in-molecules fragment-based approach. *J. Chem. Theory Comput.* **2015**, *11*, 950−961.
(51) Jovan Jose, K. V.; Raghavachari, K. Raman optical activity spectra for large molecules through molecules-in-molecules fragment-based approach. *J. Chem. Theory Comput.* **2016**, *12*, 585−594.
(52) Jose, K. V. J.; Raghavachari, K. Molecules-in-molecules fragment-based method for the calculation of chiroptical spectra of large molecules: vibrational circular dichroism and Raman optical activity spectra of alanine polypeptides. *Chirality* **2016**, *28*, 755−768.
(53) Jovan Jose, K. V.; Raghavachari, K. Molecules-in-molecules fragment-based method for the evaluation of Raman spectra of large molecules. *Mol. Phys.* **2015**, *113*, 3057−3066.
(54) Chou, C.-P.; Witek, H. A. ZZDecomposer: a graphical toolkit for analyzing the Zhang-Zhang polynomials of benzenoid structures. *Match-Commun. Math. Comp. Chem.* **2014**, *71*, 741−764.
(55) Zhang, F.; Zhang, H. A new enumeration method for Kekulé structures of hexagonal systems with forcing edges. *J. Mol. Struct.: THEOCHEM* **1995**, *331*, 255−260.
(56) Stewart, J. J. MOPAC: a semiempirical molecular orbital program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−103.
(57) Hostaš, J.; Řezáč, J.; Hobza, P. On the performance of the semiempirical quantum mechanical PM6 and PM7 methods for noncovalent interactions. *Chem. Phys. Lett.* **2013**, *568*, 161−166.
(58) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, U.SA., 2009.
(59) Zhao, Y.; Truhlar, D. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(60) Li, Q.; Noffke, B. W.; Liu, Y.; Li, L.-s. Understanding fundamental processes in carbon materials with well-defined colloidal graphene quantum dots. *Curr. Opin. Colloid Interface Sci.* **2015**, *20*, 346−353.
(61) Li, J.; Haycraft, C.; Iyengar, S. S. Hybrid extended Lagrangian, post-Hartree−Fock Born−Oppenheimer ab initio molecular dynamics using fragment-based electronic structure. *J. Chem. Theory Comput.* **2016**, *12*, 2493−2508.
(62) Li, J.; Iyengar, S. S. Ab initio molecular dynamics using recursive, spatially separated, overlapping model subsystems mixed within an ONIOM-based fragmentation energy extrapolation technique. *J. Chem. Theory Comput.* **2015**, *11*, 3978−3991.
(63) Aihara, J.-i.; Makino, M.; Ishida, T.; Dias, J. R. Analytical study of superaromaticity in cycloarenes and related coronoid hydrocarbons. *J. Phys. Chem. A* **2013**, *117*, 4688−4697.
(64) Hajgató, B.; Deleuze, M. S.; Ohno, K. Aromaticity of giant polycyclic aromatic hydrocarbons with hollow sites: super ring currents in super-rings. *Chem. - Eur. J.* **2006**, *12*, 5757−5769.
(65) Beser, U.; Kastler, M.; Maghsoumi, A.; Wagner, M.; Castiglioni, C.; Tommasini, M.; Narita, A.; Feng, X.; Müllen, K. A C216-nanographene molecule with defined cavity as extended coronoid. *J. Am. Chem. Soc.* **2016**, *138*, 4322−4325.
(66) Alonso, J. M.; Díaz-Álvarez, A. E.; Criado, A.; Pérez, D.; Peña, D.; Guitián, E. [16] Cloverphene: a clover-shaped cata-condensed nanographene with sixteen fused benzene rings. *Angew. Chem., Int. Ed.* **2012**, *51*, 173−177.
(67) Cai, J.; Ruffieux, P.; Jaafar, R.; Bieri, M.; Braun, T.; Blankenburg, S.; Muoth, M.; Seitsonen, A. P.; Saleh, M.; Feng, X.; Mullen, K.; Fasel, R. Atomically precise bottom-up fabrication of graphene nanoribbons. *Nature* **2010**, *466*, 470−473.
(68) Yan, X.; Cui, X.; Li, B.; Li, L.-s. Large, solution-processable graphene quantum dots as light absorbers for photovoltaics. *Nano Lett.* **2010**, *10*, 1869−1873.
(69) Jose, K. V. J.; Raghavachari, K. Fragment-Based Approach for the Evaluation of NMR Chemical Shifts for Large Biomolecules Incorporating the Effects of the Solvent Environment. *J. Chem. Theory Comput.* **2017**, *13*, 1147−1158.

https://dx.doi.org/10.1021/acs.jctc.9b00566
J. Chem. Theory Comput. 2020, 16, 2160−2171