JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

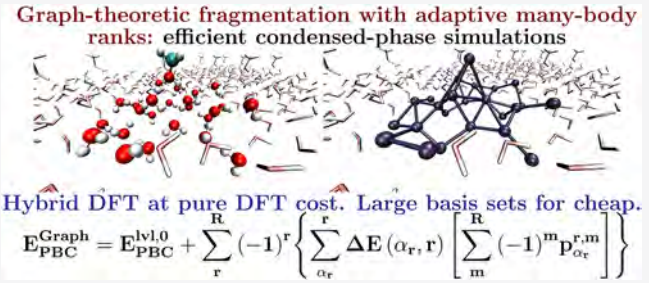
Efficient and Accurate Approach To Estimate Hybrid Functional and Large Basis-Set Contributions to Condensed-Phase Systems and Molecule−Surface Interactions

Timothy C. Ricard and Srinivasan S. Iyengar*

Cite This: J. Chem. Theory Comput. 2020, 16, 4790−4812 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We present a graph theoretic approach to adaptively compute contributions from many-body approximations in an efficient manner and perform accurate hybrid density functional theory (DFT) electronic structure calculations for condensed-phase systems. The salient features of the approach are ONIOM-like. (a) The full-system calculation is performed at a lower level of theory (pure DFT) by enforcing periodic boundary conditions. (b) This treatment is then improved through a correction term that captures many-body interactions up to any given order within a higher (in this case, hybrid DFT) level of theory. (c) In the spirit of ONIOM, contributions from the many-body approximations that arise from the higher level of theory [i.e., from (b) above] are included through extrapolation from the lower level calculation. The approach is implemented in a general, system-independent, fashion using the graph-theoretic functionalities available within Python. For example, the individual one-body components within the unit cell are designated as "nodes" within a graph. The interactions between these nodes are captured through edges, faces, tetrahedrons, and so forth, thus building a many-body interaction hierarchy. Electronic energy extrapolation contributions from all of these geometric entities are included within the above-mentioned ONIOM paradigm. The implementation of the method simultaneously uses multiple electronic structure packages. Here, for example, we present results which use both the Gaussian suite of electronic structure programs and the Quantum ESPRESSO program within a single calculation. Thus, the method integrates both plane-wave basis functions and atom-centered basis functions within a single structure calculation. The method is demonstrated for a range of condensed-phase problems for computing (i) hybrid DFT energies for condensed-phase systems at pure DFT cost and (ii) large, triple-zeta, multiply polarized, and diffuse atom-centered basis-set energies at computational costs commensurate with much smaller sets of basis functions. The methods are demonstrated through calculations performed on (a) homogeneous water surfaces as well as heterogeneous surfaces that contain organic solutes studied using two-dimensional periodic boundary conditions and (b) bulk simulations of water enforced through three-dimensional periodic boundary conditions. A range of structures are considered, and in all cases, the results are in good agreement with those obtained using large atom-centered basis and hybrid DFT calculations on the full periodic systems at significantly lower cost.



1. INTRODUCTION

Condensed-phase reactive systems are of significant interest in a wide range of chemical and material problems.[1−4] There has also been substantial effort directed toward gauging the reasons behind the relatively high efficiency of organic reactions on the surface of water droplets and interfaces.[5−10] At the core of all these important studies is the accurate determination of electronic properties for periodic condensed-phase systems.[11−16] There have been several studies[11,12,14,15,17−19] that attempt to expand the application of forefront electronic structure methods to the condensed phase. These studies include bulk systems,[20−25] surfaces,[26−28] and one-dimensional chains.[29−31] Yet, as of today, there is a disconnect between the quality of electronic structure methods and basis sets routinely available for the study of both finite-sized clusters and those available to study condensed-phase, periodic systems. For example, there exists a hierarchy of accurate density functional approaches for cluster-like studies, where hybrid functionals (the so-called rung-4 functionals[32−34]) and the more advanced double-hybrid functionals[35−37] are at the core of many successful studies.[34] However, these methods become catastrophically expensive when applied to condensed-phase systems. Furthermore, there exists a range of high-quality basis

Received: November 4, 2019
Published: June 25, 2020

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

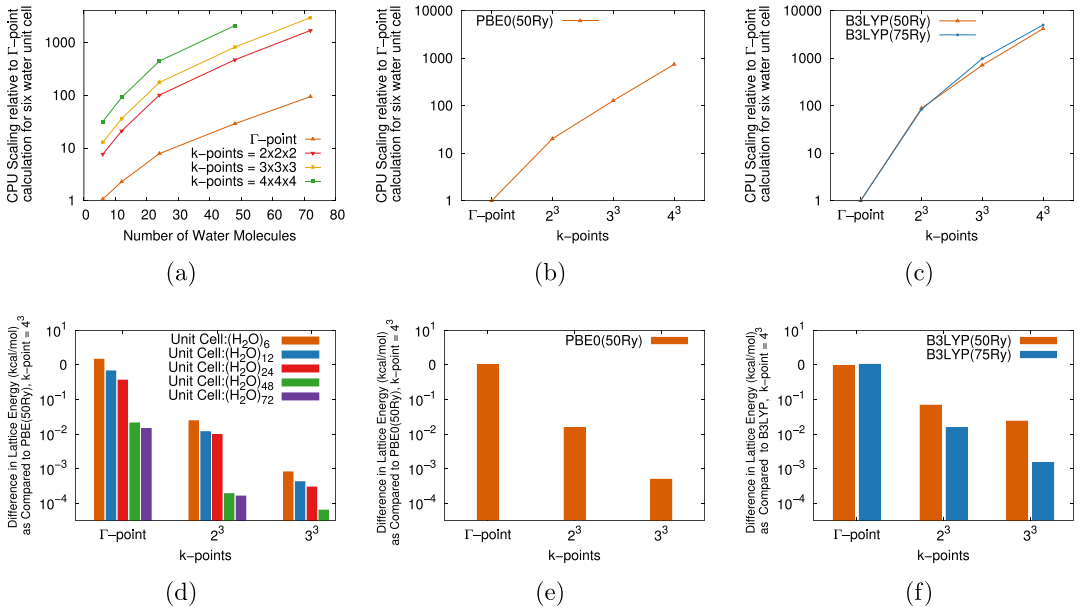


Figure 1. (a) Computational scaling for condensed-phase treatment of water using three-dimensional periodic boundary conditions: PBE functional with plane-wave basis set and a kinetic energy cutoff of 50 Rydberg using the Quantum ESPRESSO package.[45] (b) Computational scaling of the hybrid PBE0 functional, using the Kresse−Joubert style projected augmented wave treatment,[46,47] for the bulk 6 water unit cell system. (c) Computational scaling of the hybrid B3LYP functional, using a norm-conserving pseudopotential,[48,49] for the bulk 6 water unit cell system. Both for the hybrid functionals, the scaling costs, with *k*-points and unit cell size, are significantly higher than their parent functionals (see SI-1 for BLYP scaling in the plane-wave basis). (d) Convergence of the lattice energy with increasing *k*-mesh density, the effect of the *k*-points on lattice energy error decreases with the increase in unit cell size. The systems with unit cells containing 6, 12, and 24 water molecules have acceptable accuracy beyond the *k*-mesh of 2^3, whereas the 48 and 72 water unit cell systems appear to be well represented by the Γ-point mesh. Note that the 72 water unit cell results are compared against the 3^3 *k*-mesh because the 4^3 *k*-mesh is computationally too expensive to perform for the larger unit cell. (e,f) Similar to (d), but for the hybrid PBE0 and B3LYP functionals for the bulk 6 water unit cell system, corresponding to the scaling (b,c). (e,f) Similar convergence with *k*-mesh as in (d). However, the computational scaling with the increase in *k*-points is cost prohibitive for larger unit-cell sizes. It has been shown that hybrid DFT functionals demonstrate similar convergence with *k*-points as pure DFT;[50−52] thus, we use (e,f) to argue the *k*-point convergence of hybrid functionals to be the same as the pure functional in (d) for the larger unit cells. Panels (a−f) are in logarithmic scale. More details are in Supporting Information, Figures SI-1 and SI-2. The structures used in this study are described in Section 3.2 with more details found in the Supporting Information. One critical contribution from this paper is to make hybrid (and higher rung) calculations efficient and accurate for homogeneous and heterogeneous condensed-phase systems and surfaces.

sets that are routinely available for the study of isolated systems as well as systems embedded in continuum, yet these become computationally intractable for condensed-phase problems. The situation becomes worse for ab initio molecular dynamics studies[38,39] and quantum nuclear studies[40] in the condensed phase. In these cases, both the number of electronic structure calculations and quality of each such calculation affect the overall cost. In the condensed phase, this scaling problem is compounded by the use of (a) *k*-integration (reciprocal space sampling) for periodic systems, (b) diffuse and extended basis functions, and (c) nonlocal exchange−correlation functionals that are implicit within the higher rung density functionals.[12,39,41−44] As a result, accurate density functionals are cost prohibitive to use for routine periodic system studies with *k*-integration. The accurate modeling of reactivity, adsorption of substrates on surfaces, and surface defects are drastically affected by these limitations. Figure 1 briefly summarizes these challenges with additional complementary data provided in Supporting Information, Figures SI-1 and SI-2.

Apart from computational costs, atom-centered basis sets containing diffuse functions adapted to Bloch symmetry also suffer from numerical instabilities when *k*-integration[12] is included. Although homogeneous surfaces may not, in general, need large basis functions, surfaces with adsorbates and defects involve local electronic effects that are best captured through high-quality basis sets and through appropriate choice of *k*-mesh. A variety of methods have been developed to alleviate the steep scaling and numerical instabilities in periodic electronic structure calculations,[11,12,53−57] yet condensed-phase calculations always involve a trade-off between choice of basis set and computational complexity, with most production calculations conducted within the Γ-point approximation for reciprocal space sampling. While the inclusion of diffuse functions allows a better representation of weak nonbonded interactions, these calculations also often suffer from linear dependencies of basis functions and SCF convergence instabilities with accompanying significant increase in the numerical cost.

In this publication, we present an efficient approach to perform accurate electronic structure calculations for extended systems in agreement with hybrid (rung-4) functionals such as B3LYP,[58] CAM-B3LYP,[59] TPSSh,[60] and PBE0[61] at computational costs that are commensurate with semilocal (rung-2 and rung-3) functionals such as BLYP,[62] PBE,[63] and revTPSS.[64] Numerical demonstrations are conducted including both plane-wave basis functions as well as atom-centered basis functions adapted to Bloch symmetry. Furthermore, we also present results accurate to high-quality basis sets, such as 6-311++G(2df,pd) for periodic systems (i.e., adapted to Bloch symmetry), at numerical effort commensurate with moderate basis sets such as 6-31+G(d). These studies are conducted for three-dimensional bulk water problems as well as surfaces of water with adsorbates.

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

Our approach is based on ONIOM.[65−67] Here, the full system is treated using a low-level (pure density functional and/or lower quality basis) periodic calculation using either plane-wave basis functions or Bloch symmetry-adapted atom-centered Gaussian basis functions. In ONIOM-style, this calculation is corrected through "model" calculations that are now obtained from many-body expansions[68−72] computed through a graph-theoretic formalism. Thus, the "model" calculations are obtained from a graphical decomposition of a molecular cluster embedded within the full periodic system. In the studies shown here, the molecular structure contains the elements within a single unit cell, where higher level (basis set and density functional) corrections are computed based on the graphical decomposition and these perturbatively improve upon the periodic low-level calculation. In this approach, the overall energy (and gradients) for the condensed-phase system are constructed using multiple independent calculations. These include the full-system low-level periodic calculation and the molecular fragments obtained from the graphical decomposition. Our hybrid C++/Python-based computational implementation allows the use of multiple electronic structure packages for these individual segments. In this paper, we have used Gaussian[73] as well as Quantum ESPRESSO[45] for the periodic full-system calculation, where the former has allowed the use of a very large number of *k*-points; the molecular fragments that depict local many-body interactions are obtained from graphical decomposition and processed using Gaussian. In general, the computer programs developed here are able to use Gaussian, Psi4, Orca, OpenMX, and Quantum ESPRESSO simultaneously through our code for a single electronic structure, AIMD, and potential energy surface step.[74−80] Following along similar lines as ref 79, the many-body decomposition of the "model" system in the ONIOM formalism is presented using the graph theory. This idea provides the following computational advantages. The graph theory libraries in Python are extremely well refined and robust and allow for implementation of a completely general partitioning scheme. As a result, the many-body decomposition is adaptive and system independent. One-body (nodes within the graph), two-body (edges), three-body (triangles), four-body (tetrahedrons), and larger order interactions are easily constructed for any system using the appropriate set of Python libraries once the graphical representation of the molecular system is created. In this paper, we also demonstrate the convergence of our approach with increasing order of many-body terms.

Although the approach is based on ONIOM,[65−67] it is also deeply connected to other molecular fragmentation methods previously developed by our group[74−79] and by others[81−88] within the rich molecular fragmentation literature[89−98] and is also connected to many-body methods.[68−72,93,99−101] In this sense, the approach is deeply connected with the work in refs 81, 84, and 86, all of which have similar conceptual underpinnings. The effectiveness of methods involving many-body expansions (MBE)[68,93,99−102] and the related idea of molecular fragmentation[67,74−78,81−84,89−92,94,97,103−113] have been extensively demonstrated for molecular cluster calculations. These are known to remedy the intractability of correlation and extended basis-set effects especially for AIMD studies[74,75,77,78,96,114−118] and for molecular potential surfaces.[79,119] Similarly MBE and fragmentation-based methods have been employed for condensed-phase calculations.[16,18,106,120] In these methods, local quantum mechanical treatments correct long-range periodic interactions, where the latter are considered through (a) embedded charges,[106,120−129] (b) advanced molecular mechanics, using predetermined or polarizable force fields,[88,130−133] or (c), as it is done in this paper, condensed-phase calculations with HF or density functional theory (DFT) treatments.[86,134−138] Many of these methods are based upon, or closely related to, MBEs[68−72] which have been a work-horse in the study of both gas-phase clusters and condensed-phase systems.

We demonstrate our method for thin films of water/ice, considered because of potential catalytic significance.[5,6,8,10,139,140] We consider a range of molecular geometries for each case, and these include both stable (equilibrium) structures as well as those that may be encountered during dynamics studies (and may hence be farther from equilibrium) to gauge the accuracy of the proposed methods. At the end, we find the proposed approach to be capable of providing energies in agreement with hybrid functionals (rung-4) and large atom-centered basis-set calculations, at the accuracy of the order of fractions of kcal/mol at much reduced computational cost.

The paper is organized as follows: In Section 2, we present our graph-theoretic approach to low-cost condensed-phase calculations. In Section 3.2, we discuss the systems used in our benchmarks, and these are further expounded upon in Supporting Information Section SI-1. Many-body or graph-theoretic rank dependence of accuracy is discussed in Section 4.1 and further substantiated by the Supporting Information. The results from our benchmark studies are presented in Section 5 and again complemented by the Supporting Information. Conclusions are given in Section 6.

2. LOW-COST, ACCURATE TREATMENT OF CONDENSED-PHASE ELECTRONIC STRUCTURES THROUGH GRAPH-THEORETIC DECOMPOSITION OF MOLECULAR STRUCTURES

In recent years, molecular fragmentation methods[81−85,87,89−91,93−96,99−101,104,106,107,110,113,138,141−149] and generalized approaches to the well-known many-body theory[68,69,71,72,93,99−102] have emerged as powerful alternatives to alleviate the critical scaling problem of post-Hartree−Fock methods and large basis-set calculations. There have been several efforts where the long-range effects in condensed-phase systems are treated in an empirical fashion, with the real-space, local finite portions of the full system being treated using higher levels of theory.[120−128,132,133] More recently there has been an increased focus on fragmentation studies that have included nonempirical treatments of long-range effects in the condensed phase to obtain post-Hartree−Fock or hybrid DFT levels of accuracy.[86,134−138] To allow accurate and efficient hybrid DFT simulations of the same, the central contribution of this paper is the introduction of long-range, condensed-phase, periodic interactions at a lower level of theory (such as semilocal, rung-2, DFT functionals using smaller basis sets), with the perturbative introduction of higher levels of theory (rung-4, hybrid density functionals) and larger basis sets using local MBEs that are obtained from graph theory. In this manner, local defects, reactive effects, and adsorption properties will be treated at the necessary higher level of electronic structure and larger basis, whereas long-range periodic effects are considered through an ONIOM-like extrapolation procedure. The approach is benchmarked later in the publication for both homogeneous (bulk water and water surfaces) as well as heterogeneous (surfaces with adsorbates) systems, where we expect the latter to be particularly challenging tests for our model. Hence, to use the relevant ONIOM terminology, the

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

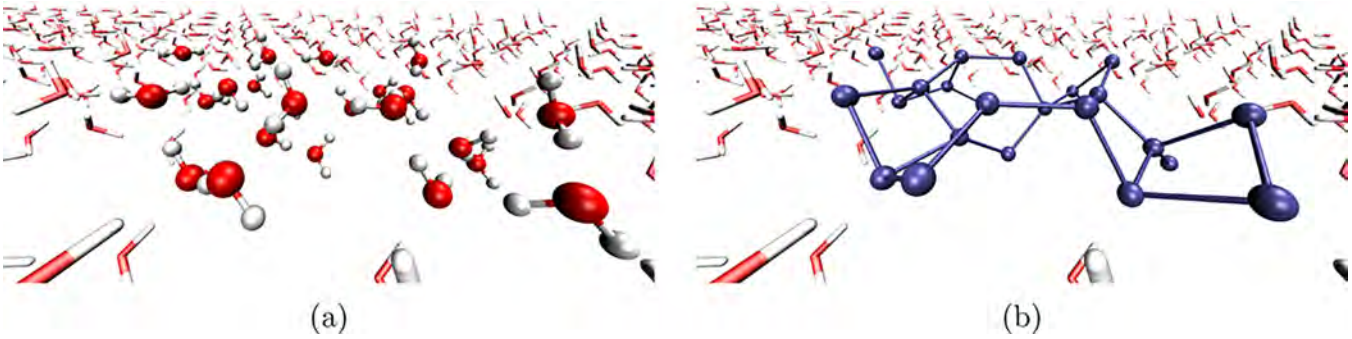


Figure 2. Basic idea of the method arising from eq 1 (and later from eqs 6 and 14) is illustrated through a periodic slab of water, with a highlighted cluster. The periodic nature of the problem is captured by $E_{\mathrm{PBC}}^{\mathrm{level,0}}$. Additional correlation, exchange, and large basis-set effects are perturbatively included for the highlighted subsystem through E_χ defined in eq 5 using the graphical network depiction of (a) in (b). The systems considered in this paper are discretized into nodes which represent single water and adsorbate molecules.

"model" perturbation to the "real" calculation is obtained using a graph-theoretic tool to efficiently include many-body interactions up to any given order. Thus, the central approximation in this work may be captured by a perturbative correction to the overall energy of a periodic system that is given by

$$E_{\mathrm{PBC}}^{\text{graph-theoretic}} = E_{\mathrm{PBC}}^{\mathrm{level,0}} + E_\chi \tag{1}$$

Here, $E_{\mathrm{PBC}}^{\mathrm{level,0}}$ refers to the periodic treatment of the system alluded to above, at some lower level of theory and basis together referred to as "level, 0". The quantity E_χ is an ONIOM-like correction term[65−67] obtained here from a graph theoretic decomposition of the molecular structure that, as we will show, captures the accuracy at higher levels of theory and basis sets. In this paper, the region of chemical interest is discretized as a set of geometric nodes that represent localized molecular fragments. These nodes are then connected to form a network or a graph. The connections between nodes, represented by edges, faces, and other higher-order objects referred to as simplexes in the graph, capture local (bonded and nonbonded) as well as nonlocal (nonbonded) many-body interactions. Such a geometrical decomposition is shown in Figure 2. Furthermore, the graph in Figure 2 has an associated topological invariance known as an Euler characteristic,[150] χ, defined below in eq 2, and guides our definition of E_χ in eq 1

$$\begin{aligned}\chi &= \eta_0 - \eta_1 + \eta_2 - +(-1)^r\eta_r + \cdots+(-1)^{\mathcal{R}}\eta_{\mathcal{R}} \\ &= \sum_{r=0}^{\mathcal{R}} (-1)^r \eta_r\end{aligned} \tag{2}$$

Here, η_r represents the number of geometric entities (nodes, edges, faces, and higher-order simplexes) of rank-r. That is, η_0 is the number of nodes (rank-0 simplexes) in the graph, η_1 is the number of edges (rank-1 simplexes) in the graph, η_2 is the number of faces (rank-2 simplexes) in the graph, and so on.[150−152] The quantity $\mathcal{R}$ is the largest simplex rank included in the truncated expansion on the right side of eq 2, which in general can go up to arbitrary orders.

In this publication, we replace the appearance of each rank-r simplex in eq 2, numbered using the index α, by an energy correction corresponding to the molecular interactions captured by the specific simplex, and thus

$$\eta_r \rightarrow \left\{ \sum_{\alpha}^{r\text{-rank}} \Delta E(\alpha, r) \left[\sum_{m=r}^{\mathcal{R}} (-1)^m p_\alpha^{r,m} \right] \right\} \tag{3}$$

where the summation over α is over all rank-r simplexes and $\Delta E(\alpha,r)$ is an ONIOM-like energy correction or extrapolation term added to $E_{\mathrm{PBC}}^{\mathrm{level,0}}$ (see eq 1) for the α-th rank-r simplex

$$\Delta E(\alpha, r) = E^{\mathrm{level,1}}(\alpha, r) - E^{\mathrm{level,0}}(\alpha, r) \tag{4}$$

Thus, in ONIOM-style, the electronic energy for the α-th rank-r simplex (representing a molecular fragment in the system) is computed at two levels of theory, namely, "level, 1" and "level, 0". (Note that the periodic calculation in eq 1 is performed at "level, 0".) The square bracketed term in eq 3 contains an overcounting correction.[77,78] The overcounting correction, $p_\alpha^{r,m}$, refers to the number of times the αth rank-r simplex appears in all rank-m simplexes ($m \geq r$), with phase $(-1)^m$. Correspondingly, in eq 1, the energy correction E_χ represents the accumulation of the measures for each simplex. In light of the topological invariant, χ, from eq 2 and the map in eq 3, the expression for E_χ is

$$E_\chi \equiv \sum_{r=0}^{\mathcal{R}} (-1)^r \left\{ \sum_{\alpha}^{r\text{-rank}} \Delta E(\alpha, r) \left[\sum_{m=r}^{\mathcal{R}} (-1)^m p_\alpha^{r,m} \right] \right\} \tag{5}$$

The curly bracketed term, { }, refers to the energetic replacement for each rank-r simplex, that is, η_r in eq 2, as per eq 3. In Section 2.1, we show that E_χ can also be regarded as a difference in many-body expansions. As it is shown, the overcounting correction, $p_\alpha^{r,m}$, is along similar lines as those present in many-body and double many-body expansions[68,93,99−101] and molecular fragmentation.[67,82,84,89,91,94,95,101,104,110,116,146] Within the many-body paradigm, $p_\alpha^{r,m}$ is the count of the number of times the αth $(r+1)$-body term appears in all $(m+1)$-body terms.

When eq 5 expression is incorporated into the approximation of a periodic system energy given by eq 1, the graph-theoretic energy expression[77] for a periodic system may be written as

$$\begin{aligned}E_{\mathrm{PBC}}^{\text{graph-theoretic}} = E_{\mathrm{PBC}}^{\mathrm{level,0}} + \sum_{r=0}^{\mathcal{R}} (-1)^r \Bigg\{ \sum_{\alpha}^{r\text{-rank}} \Delta E(\alpha, r) \\ \left[\sum_{m=r}^{\mathcal{R}} (-1)^m p_\alpha^{r,m} \right] \Bigg\}\end{aligned} \tag{6}$$

This expression, as one can see, is deeply influenced by ONIOM where the "model" system now has a more complex structure as compared to the traditional model. However, as discussed in ref 79, this structure also provides the basis to obtain

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

general potential energy surfaces where the graphs change (and the corresponding topological properties in eq 3 change) as nuclei evolve.

As noted above, "α" represents a simplex[151,152] of rank-r and $\mathcal{R}$ is the largest simplex rank (maximum order of many-body interactions as we see below) considered for electronic structure treatment. The availability of larger rank simplexes is highly dependent on the topology of the graph representation of the system. The number of available edges surrounding any given node in turn depends on the local proximity of these groups; this proximity is governed by a distance cutoff which is discussed in Section 4. The mutual proximity of these nodes influences the extent to which, as we will see in the next subsection, many-body approximations contribute to E_χ. Thus, in Section 4.1, the choice of $\mathcal{R}$ is examined with respect to accuracy of eq 6, with more details in the Supporting Information. These studies in Sections 4.1 and SI-2 are further strengthened by varying the allowed edge lengths (or two-body interactions as we will see in Section 2.1 below) within the graph in conjunction with the maximum simplex rank (or many-body interaction, as will be seen from Section 2.1).

2.1. Relation between Equation 6 and Standard Many-Body Expansions. One may clearly see the connections between eq 6 and the well-known many-body expansions (MBE)[68,93,99−101] by simply writing out the appropriate form of eq 6 for $\mathcal{R} = 1$

$$E^{\text{graph-theoretic}}_{\text{PBC},\mathcal{R}=1} = E^{\text{level},0}_{\text{PBC}} + \sum_{\alpha\in\text{edges}} \Delta E(\alpha, 1) - \sum_{\alpha\in\text{nodes}} \Delta E(\alpha, 0)[p_\alpha^{0,1} - p_\alpha^{0,0}] \tag{7}$$

where, now, $p_\alpha^{0,1}$ is the number of times the node α (one-body term) appears in all edges (or two-body interactions) and $p_\alpha^{0,0}$ is the number of times node α appears in all nodes and by extension $p_\alpha^{0,0} = 1$. Using eq 4, we may further rewrite eq 7 as

$$E^{\text{graph-theoretic}}_{\text{PBC},\mathcal{R}=1} = E^{\text{level},0}_{\text{PBC}} + \sum_{\alpha\in\text{edges}} [E^{\text{level},1}(\alpha, 1) - E^{\text{level},0}(\alpha, 1)] - \sum_{\alpha\in\text{nodes}} [E^{\text{level},1}(\alpha, 0) - E^{\text{level},0}(\alpha, 0)](p_\alpha^{0,1} - 1) \tag{8}$$

$$\begin{aligned} &= E^{\text{level},0}_{\text{PBC}} \\ &+ \left\{ \sum_{\alpha\in\text{nodes}} E^{\text{level},1}(\alpha, 0) + \left[\sum_{\alpha\in\text{edges}} E^{\text{level},1}(\alpha, 1) - \sum_{\alpha\in\text{nodes}} p_\alpha^{0,1} E^{\text{level},1}(\alpha, 0) \right] \right\} \\ &- \left\{ \sum_{\alpha\in\text{nodes}} E^{\text{level},0}(\alpha, 0) + \left[\sum_{\alpha\in\text{edges}} E^{\text{level},0}(\alpha, 1) - \sum_{\alpha\in\text{nodes}} p_\alpha^{0,1} E^{\text{level},0}(\alpha, 0) \right] \right\} \end{aligned} \tag{9}$$

Here, we have grouped terms belonging to level, 1 and level, 0 separately. The terms inside each of the curly brackets, { }, are the level, 1 (or level, 0) two-body energies. These are specifically written as a one-body term which sums over all nodes

$$E^{\text{level},1}_{\text{1-body}} \equiv \sum_{\alpha\in\text{nodes}} E^{\text{level},1}(\alpha, 0) \tag{10}$$

and similarly for level, 0. The two-body correction term is captured within the square-bracketed terms, [ ], in eq 9, having the form

$$E^{\text{level},1}_{\text{2-body,corr.}} \equiv \sum_{\alpha\in\text{edges}} E^{\text{level},1}(\alpha, 1) - \sum_{\alpha\in\text{nodes}} p_\alpha^{0,1} E^{\text{level},1}(\alpha, 0) \tag{11}$$

Thus, the two-body correction term has the form of energy contribution arising from two-body calculations (edges) minus each one-body contributions times number of times the one-body term (node) appears inside all two-body terms, ($p_\alpha^{0,1}$). Thus, we may rewrite the expression above as

$$E^{\text{graph-theoretic}}_{\text{PBC},\mathcal{R}=1} = E^{\text{level},0}_{\text{PBC}} + \{E^{\text{level},1}_{\text{1-body}} + E^{\text{level},1}_{\text{2-body,corr.}}\} - \{E^{\text{level},0}_{\text{1-body}} + E^{\text{level},0}_{\text{2-body,corr.}}\} \tag{12}$$

It is clear that the ONIOM-style extrapolation now applies in eq 12 in a many-body context. However, eq 6 includes many-body contributions to arbitrary orders. This can be seen by constructing a similar n-body analysis for eq 6, and for three-body interactions (faces), one may use $\mathcal{R} = 2$ in eq 6 to obtain

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

$$E^{\text{graph-theoretic}}_{\text{PBC},\mathcal{R}=2} = E^{\text{level},0}_{\text{PBC}} + \sum_{r=0}^{\mathcal{R}=2} (-1)^r \left\{ \sum_{\alpha}^{r\text{-rank}} \Delta E(\alpha, r) \left[\sum_{m=r}^{\mathcal{R}=2} (-1)^m p_\alpha^{r,m} \right] \right\} = E^{\text{level},0}_{\text{PBC}}$$
$$+ \left\{ \sum_{\alpha\in\text{nodes}} E^{\text{level},1}(\alpha, 0) + \left[\sum_{\alpha\in\text{edges}} E^{\text{level},1}(\alpha, 1) - \sum_{\alpha\in\text{nodes}} p_\alpha^{0,1} E^{\text{level},1}(\alpha, 0) \right] \right.$$
$$\left. + \left[\sum_{\alpha\in\text{faces}} E^{\text{level},1}(\alpha, 2) - \sum_{\alpha\in\text{edges}} p_\alpha^{1,2} E^{\text{level},1}(\alpha, 1) + \sum_{\alpha\in\text{nodes}} p_\alpha^{0,2} E^{\text{level},1}(\alpha, 0) \right] \right\}$$
$$- \left\{ \sum_{\alpha\in\text{nodes}} E^{\text{level},0}(\alpha, 0) + \left[\sum_{\alpha\in\text{edges}} E^{\text{level},0}(\alpha, 1) - \sum_{\alpha\in\text{nodes}} p_\alpha^{0,1} E^{\text{level},0}(\alpha, 0) \right] \right.$$
$$\left. + \left[\sum_{\alpha\in\text{faces}} E^{\text{level},0}(\alpha, 2) - \sum_{\alpha\in\text{edges}} p_\alpha^{1,2} E^{\text{level},0}(\alpha, 1) + \sum_{\alpha\in\text{nodes}} p_\alpha^{0,2} E^{\text{level},0}(\alpha, 0) \right] \right\}$$
$$= E^{\text{level},0}_{\text{PBC}} + \{E^{\text{level},1}_{\text{1-body}} + E^{\text{level},1}_{\text{2-body,corr.}} + E^{\text{level},1}_{\text{3-body,corr.}}\} - \{E^{\text{level},0}_{\text{1-body}} + E^{\text{level},0}_{\text{2-body,corr.}} + E^{\text{level},0}_{\text{3-body,corr.}}\} \quad (13)$$

It is clear from eqs 12 and 13 that eq 6 provides an ONIOM-type extrapolation approximation (" high" minus "low" or level, 1 minus level, 0) where the "high" and "low" levels are themselves many-body approximations

$$E^{\text{graph-theoretic}}_{\text{PBC},\mathcal{R}} = E^{\text{level},0}_{\text{PBC}} + E^{\text{level},1}_{\text{MBE},\mathcal{R}} - E^{\text{level},0}_{\text{MBE},\mathcal{R}} \quad (14)$$

where $E^{\text{level},1}_{\text{MBE},\mathcal{R}}$ (and similarly $E^{\text{level},0}_{\text{MBE},\mathcal{R}}$) is many-body expansion of order $\mathcal{R}$. Thus, eq 6 (and correspondingly eq 14) provides an adaptive recipe to compute interactions to arbitrary orders through an efficient graph-theoretic decomposition. The generality of the approach is strongly dependent on the Python implementation of eqs 6 and 14, which provides a robust medium containing a rich set of functionalities for graph-theoretic studies.

A critical reason why eq 14 works out from eq 6 is because for higher rank simplexes (up to rank $\mathcal{R}$) our graph-theoretic formalism is strongly influenced by the fact that simplexes are closed convex hulls.[151,152] Thus, a higher rank simplex is one that includes within it all component, lower-rank, simplexes. For instance, a rank-2 simplex, a triangle, contains all three edges, that is, rank-1 simplexes (consistent with the envelope protocols above) and a rank-3 simplex, tetrahedron, includes within all four of its component triangles (rank-2 simplexes). These prescriptions directly follow from the fact that simplexes are closed convex hulls and the edges of any rank-r simplex are affinely independent.[153] Thus, algebraically speaking, a rank-r simplex, is constructed from the family linearly independent rank-1 simplexes (edges), $\{u_i\}$ as

$$S_r = \sum_{i=0}^{r} \lambda_i u_i \quad (15)$$

where $\sum_{i=0}^{r} \lambda_i = 1$ and $\lambda_i \geq 0$. Therefore, higher-order terms (up to rank $\mathcal{R}$) are automatically included when the monomers are mutual "neighbors" within the graph. This implies that the choice of the spatial envelope is a means of controlling the maximum rank object considered. This requirement of simplexes being convex hulls is the single main constraint in the algorithm that maps our approach to many-body theory as seen above. The absence of this computational restriction will allow including higher-order many-body interactions that do not then properly cancel the lower-body contributions and are inconsistent with the set-theoretic inclusion exclusion principle.[74,154] Furthermore, eqs 6 and 14, due to their ONIOM underpinnings, introduce a difference in rank-$\mathcal{R}$ many-body contributions computed from two levels of theory and basis. Through this approach, the higher rank many-body interactions are already included through $E^{\text{level},0}_{\text{PBC}}$. It must be noted that this is also the essence of the HMBI method from Beran and co-workers,[147] where the supersystem many-body terms beyond a certain order are either treated using polarizable force-fields[131,155−159] or more recently using Hartree−Fock.[86,129,160]

Here, we have presented the application of our graph-theoretic interpretation[77−79] in eqs 6 and 14 to reduce the costs and instabilities in condensed-phase calculations, while presenting a completely adaptive scheme that is suited for ab initio dynamics and computing molecular potential surfaces as evidenced by refs 74 and 79. As noted in eq 6, this is done by capturing local interactions incorporated into the graphical representation (in this case the unit cell, see Figure 2a,b) with large basis sets containing higher-order diffuse and polarized basis functions, while modeling the bulk properties with lower quality, computationally less expensive basis sets. Similarly, we also captured local many-body interactions within the molecular fragments with hybrid and other higher rung DFT functionals, while modeling the long-range, periodic interactions with the more localized "pure" functionals. For both condensed-phase and cluster calculations, the many-body effects are already treated at a base electronic structure level, and only corrections from this base level are computed using the difference in eq 5 to obtain eq 14. When considering condensed-phase systems, local interactions between the constituents of the unit cell and the periodic, condensed phase are both critical aspects to understanding the physics of the system.

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

Table 1. Systems and Configurations for Condensed Systems with Two- and Three-Dimensional Periodic Boundary Conditions[a]

unit cell composition	condensed-phase system	number of structures	energy[b] spread	mean[c] deviation	$k_{\text{graph-theoretic}}$[d]	$k_{\text{hybrid-DFT}}$[e]
		Optimized at BLYP/6-31+G(d)				
$(H_2O)_6$	gas phase	15	3.7	7.7		
$(H_2O)_{12}$	2D surface	42	10.8	22.1	71 ± 28	285 ± 108
$(H_2O)_{24}$	2D surface	19	19.6	30.3	43 ± 17	191 ± 54
$(H_2O)_6$	3D bulk	15	4.8	9.8	228 ± 46	5944 ± 668
		Optimized at PBE[f] (50 Ry)				
$(H_2O)_6$	3D bulk	15	4.7	9.3	8	8
$(H_2O)_{12}$	3D bulk	30	9.4	19.0	8	8
$(H_2O)_{24}$	3D bulk	32	18.0	37.0	1	1
$(H_2O)_{48}$	3D bulk	14 (24)[g]	20.1	66.2	1	1

[a]More details can be found in Supporting Information, Figures SI-5, 6, 8 and Table SI-I with discussion in Section SI-I. [b]Standard deviation of the electronic energies (in kcal/mol) at BLYP/6-31+G(d) for the structures in the top block of data. Bottom block: structures optimized with PBE with the Kresse–Joubert style pseudopotential[47] and 50 Ry cutoff for plane waves.[45] Energy distribution of these structures is shown in detail in Figure SI-5, 6, and 8. [c]Deviation of the mean energy of the conformational population from the lowest energy configuration within the respective datasets. A detailed distribution is provided in Figures SI-5, 6, and 8. [d]The quantity $k_{\text{graph-theoretic}}$ refers to the number of k-space sampling points needed to compute $E_{\text{PBC}}^{\text{level},0}$ in eq 6. In our study, semilocal DFT functionals were used for level, 0 calculations, to extrapolate to level, 1 that was generally hybrid DFT. Hence, in these cases, the choice of semilocal DFT functionals dictate the number of k-points. Average and Standard deviations are presented. [e]Number of k-space sampling points needed during the higher level periodic calculations in Section 5. Average and standard deviations are presented. [f]Projector-augmented wave (PAW) representation[46] with the Kresse–Joubert style pseudopotential.[47] [g]Number of structures sampled from AIMD simulations are shown within parenthesis. The number outside parenthesis is the number obtained from geometry optimization. See Figure SI-6e.

Table 2. Systems and Configurations Considered for Adsorbate Systems

unit cell composition	condensed-phasesystem	number of structures	energy[a] spread	mean[b] deviation	$k_{\text{graph-theoretic}}$[c]	$k_{\text{hybrid-DFT}}$[d]
$(H_2O)_{24}$·H_2	2D surface with adsorbate	39 (13)[e]	35.6	53.9	33 ± 15	161 ± 57
$(H_2O)_{24}$·CH_4	2D surface with adsorbate	44 (13)[e]	28.7	52.5	32 ± 14	166 ± 48
$(H_2O)_{24}$·CH_3OH	2D surface with adsorbate	32 (13)[e]	29.1	57.0	27 ± 7	154 ± 33

[a]Standard deviation of the electronic energies at revTPSS/6-31+G(d) (in kcal/mol). [b]Deviation of the mean energy from the lowest energy configuration within the respective data-sets, at revTPSS/6-31+G(d) (in kcal/mol). [c]Number of k-space sampling points needed for $E_{\text{PBC}}^{\text{level},0}$. In our study, semilocal DFT functionals were used for level, 0 and hence dictate the number of k-points in this case. Average and standard deviations are presented. [d]Number of k-space sampling points needed during the higher level periodic calculations in Section 5. Average and standard deviations are noted. [e]Number of structures sampled from AIMD are shown within parenthesis. The number outside parenthesis is the number obtained from geometry optimization. See Figure SI-12.

3. CONDENSED-PHASE SYSTEMS CONSIDERED AND THE ASSOCIATED DISCUSSION OF RANK DEPENDENCE ON ACCURACY

The systems considered in this paper are both homogeneous as well as more challenging heterogeneous systems. These include three-dimensional periodic bulk water, two-dimensional periodic thin films of water, and adsorbate molecules, H_2, methane, and methanol bound to the surface of water films. The accuracy and efficiency of the aforementioned graph-theoretic approximations from eq 6 are gauged toward computation of higher rung DFT and large, diffused, polarized, and split-valence atom-centered basis-set description of condensed-phase problems. The latter may especially be important for the surface adsorption studies.

In Section 3.1, we discuss the computational implementation of our studies and also clarify the notation used in the paper. Section 3.2 introduces the library of structures used in the benchmarks with more details in the Supporting Information.

3.1. Computational Aspects. Our implementation of eq 6, through an MPI parallelized hybrid C++/Python module, is capable of using multiple electronic structure packages within a single AIMD step and isolated electronic structure calculations conducted during potential surface evaluations.[79] The code currently supports the following set of external electronic structure packages during a single energy and gradient evaluation: Gaussian,[73] ORCA,[161] Psi4,[162] Quantum ESPRESSO,[45] and OpenMX.[163]

The main components of eq 6, the full-system low-level energy, $E_{\text{PBC}}^{\text{level},0}$, and the fragment energies, $\{E^{\text{level},1}(\alpha,r); E^{\text{level},0}(\alpha,r)\}$, are independent and can be performed simultaneously without constraints on the electronic structure package used. This is precisely what is done here, and in the analogous molecular cluster studies in refs 77 and 78. In this publication, we have used both the Gaussian series of electronic structure programs[73] (using atom-centered basis sets, with and without Bloch symmetry enforced for periodic boundary conditions) and Quantum ESPRESSO[45] (using plane waves with pseudopotentials) to compute the various energy terms in eq 6. The low-level full-system energies, $E_{\text{PBC}}^{\text{level},0}$, are computed here using both electronic structure packages. This allows the detailed evaluation of our method for applications using both plane-wave basis functions (with pseudopotentials) and all-electron atom-centered basis functions adapted to Bloch symmetry.[12] The fragment energies, $\{E^{\text{level},1}(\alpha,r); E^{\text{level},0}(\alpha,r)\}$, in all cases are computed here using Gaussian-basis cluster calculations. Specifically, we demonstrate the implementation of eq 6 with the condensed-phase contribution, $E_{\text{PBC}}^{\text{level},0}$, treated with pure DFT using either atom-centered[73] and using plane-wave[45] basis sets. As required by eq 6, these energies are then corrected by the differences in local many-body interactions (see eq 14) that are obtained from atom-centered basis function calcu-

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

lations. Thus, in part, we also provide an approach to correct plane-wave results with higher level of theory, including higher rung DFT treatment.

Ab initio treatment of periodic systems requires an appropriate choice of k-point sampling. The k-mesh chosen for the atom-centered basis treatment of each system was automatically generated as part of the Gaussian electronic structure calculations.[73] This was done based on the relative size of the reciprocal space, thus leading to fewer required k-points with the increase in lengths of the respective translation vectors.[11,164] In contrast, when using Quantum ESPRESSO, the number of k-points was determined by the lattice energy convergence tests in Figure 1. Additionally, our calculations with atom-centered bases in the condensed phase used the methods described in refs 11 and 12 which allowed the expansion of k-point mesh without significantly increasing the overall computational costs. The k-point sampling chosen for each of the set of systems considered is shown in Tables 1 and 2 below. It is also clear from Table 1 that for the use of atom-centered basis functions there are far fewer number of k-points necessary in energies with pure DFT functionals as compared to the case for hybrid functionals. In addition, the use of Bloch symmetrized atom-centered basis functions, to compute $E_{\mathrm{PBC}}^{\mathrm{level},0}$, allows the flexibility to not require periodic boundary conditions to be enforced in all directions, which allows for a lower dimensional k-mesh. This is especially important for surface adsorption problems, as the addition of periodic boundary and k-integration through the vacuum leads to significant computational costs. Furthermore, we have used atom-centered diffuse basis functions in periodic calculations. In most condensed matter studies, smaller basis sets are generally used because (a) the cost of using diffuse functions can be prohibitive and (b) the periodic nature of the system may yield a situation where basis functions get automatically delocalized through Bloch symmeterization. But in this paper, because the goal was to make the more challenging calculations (larger basis and hybrid calculations) easier to perform, we have decided to use diffuse functions in all our periodic treatments.

The use of eq 6 implies that the extended periodic system is treated at a choice of theory and basis (level, 0, leading to $E_{\mathrm{PBC}}^{\mathrm{level},0}$) that bares acceptable cost and numerical stability for the chosen systems. Here, we chose DFT with semilocal exchange functionals with minimal diffuse and polarized atom-centered basis functions such as PBE/6-31+G(d) or plane-wave basis with pseudopotentials[44,46−49] such as PBE with a 50 Rydberg cutoff and Kresse−Joubert style pseudopotential.[47] The perturbative correction of eq 6 lead to simplexes that are treated at level, 1 and level, 0. This aspect may also be clear from eq 14 and the discussion above. The desired accuracy (level, 1) was chosen to be (a) hybrid density functionals with a plane-wave or modest atom-centered basis (Section 5.1) and (b) semilocal DFT and screened Coulomb hybrid GGA, density functionals with large basis treatments (Section 5.2). Thus, we present efficient, accurate, and stable[165] periodic calculations at higher levels of density functional and basis-set approximations.

Considering that eq 6 is computed here using multiple levels of theory, we introduce a notation to ease the discussion for the remaining portion of the paper. This notation is slightly different based on whether Quantum ESPRESSO (plane waves) or Gaussian (atom-centered Gaussian basis adapted to Bloch symmetry) is used to compute $E_{\mathrm{PBC}}^{\mathrm{level},0}$. When the atom-centered basis functions are used, the notation "PBE0:PBE/6-31+G(d)" implies that the $E_{\mathrm{PBC}}^{\mathrm{level},0}$ and $E^{\mathrm{level},0}(\alpha,r)$ terms in eqs 6 and 4 are computed at the PBE/6-31+G(d) level while the target level of theory is PBE0/6-31+G(d). Furthermore, $E^{\mathrm{level},1}(\alpha,r)$ in eqs 6 and 4 is obtained from PBE0/6-31+G(d). In this case, both the target level of theory and $E_{\mathrm{PBC}}^{\mathrm{level},0}$ use Bloch-symmetry adapted forms of the basis, whereas this is not the case for $E^{\mathrm{level},0}(\alpha,r)$ and $E^{\mathrm{level},1}(\alpha,r)$. Similarly, the notation "PBE/6-311++G(2df,pd):6-31+G(d)" implies that PBE/6-31+G(d) is used in computing $E_{\mathrm{PBC}}^{\mathrm{level},0}$ and $E^{\mathrm{level},0}(\alpha,r)$, whereas PBE/6-311++G(2df,pd) is used to obtain $E^{\mathrm{level},1}(\alpha,r)$. Again, target and $E_{\mathrm{PBC}}^{\mathrm{level},0}$ use Bloch-symmeterized forms. When $E_{\mathrm{PBC}}^{\mathrm{level},0}$ is computed using a plane-wave basis from Quantum ESPRESSO, the DFT functional and kinetic energy cutoff employed are appended to the nomenclature above. For example, "PBE0:PBE/6-31+G(d):PBE(50 Ry)" indicates that $E_{\mathrm{PBC}}^{\mathrm{level},0}$ is treated with PBE using a plane-wave basis that has a cutoff of 50 Rydberg. In addition, the terms $E^{\mathrm{level},0}(\alpha,r)$ and $E^{\mathrm{level},1}(\alpha,r)$ are computed using atom-centered basis functions, using the PBE/6-31+G(d) and PBE0/6-31+G(d) levels of theory, respectively. The target level of theory for these calculations is PBE0(50 Ry). Thus, all notations above may be summarized by including both extrapolation and target levels as "PBE0:PBE/6-31+G(d) → PBE0/6-31+G(d)(Bloch)", "PBE/6-311++G(2df,pd):6-31+G(d) → PBE/6-311++G(2df,pd)(Bloch)", and "PBE0:PBE/6-31+G(d):PBE(50 Ry) → PBE0(50 Ry)".

The accuracy of this method is quantified from errors in lattice energy across a variety of structures discussed in Section 3.2. The lattice energy per monomer is defined as

$$E_{\mathrm{lattice}} = \frac{1}{N_{\mathrm{monomer}}}\left(E_{\mathrm{PBC}}(i) - \sum_{j=1}^{N_{\mathrm{monomer}}} E_{\mathrm{monomer}}(j)\right) \tag{16}$$

where the summation is included so as to also represent the case for nonhomogenous adsorbate systems. We compute errors for both the energies obtained from eq 6 and from using the lower level of theory as an approximation to full periodic calculations at higher levels. These errors are computed according to

$$\epsilon_{\mathrm{Lattice}} = \frac{1}{N_{\mathrm{structures}}}\left(\sum_{i}^{N_{\mathrm{structures}}} |E_{\mathrm{lattice}}^{\mathrm{level},1} - E_{\mathrm{lattice}}^{\mathrm{b}}|\right) \tag{17}$$

The reported error in lattice energy ($\epsilon_{\mathrm{Lattice}}$) is the mean absolute difference between the lattice energies for the full periodic target treatment, $E_{\mathrm{PBC}}^{\mathrm{level},1}$, and the method referred to as "b" in eq 17, which would be either the graphical extrapolation based on eq 6, that is, $E_{\mathrm{PBC,R}}^{\mathrm{graph\text{-}theor.}}$, or the baseline full periodic low-level treatment $E_{\mathrm{PBC}}^{\mathrm{level},0}$. The latter is used in the benchmark studies to gauge the improvement afforded from using $E_{\mathrm{PBC}}^{\mathrm{graph\text{-}theor.}}$ versus the use of only the low-level periodic treatment, $E_{\mathrm{PBC}}^{\mathrm{level},0}$, to model $E_{\mathrm{PBC}}^{\mathrm{level},1}$.

It may also be noted that in eq 17, the monomer energies cancel out when the same electronic structure package is used for both fragmentation ($E_{\mathrm{PBC,R}}^{\mathrm{graph\text{-}theor.}}$) and the reference level, 1 calculation ($E_{\mathrm{PBC}}^{\mathrm{level},1}$). In that case, eq 17 is a direct gauge of the absolute energy differences between $E_{\mathrm{PBC,R}}^{\mathrm{graph\text{-}theor.}}$ and $E_{\mathrm{PBC}}^{\mathrm{level},1}$. For the case where $E_{\mathrm{PBC}}^{\mathrm{level},1}$ is computed with Quantum ESPRESSO (PBE0) and $E_{\mathrm{PBC,R}}^{\mathrm{graph\text{-}theor.}}$ using both Quantum ESPRESSO ($E_{\mathrm{PBC}}^{\mathrm{level},0}$, pure functional) and Gaussian (fragments), the comparison is more complicated. However, it may also be noted from Figures SI-6 and SI-8 that the conformational energies span an energetic landscape of the order of 1−10 kcal/mol. As we will see later in the discussion (Sections 4 and 5), the errors from eq 17 noted in all cases are of the order of 0.1 kcal/mol or less. This implies that the energetic ordering of conformers in Figures SI-6 and SI-8, as

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

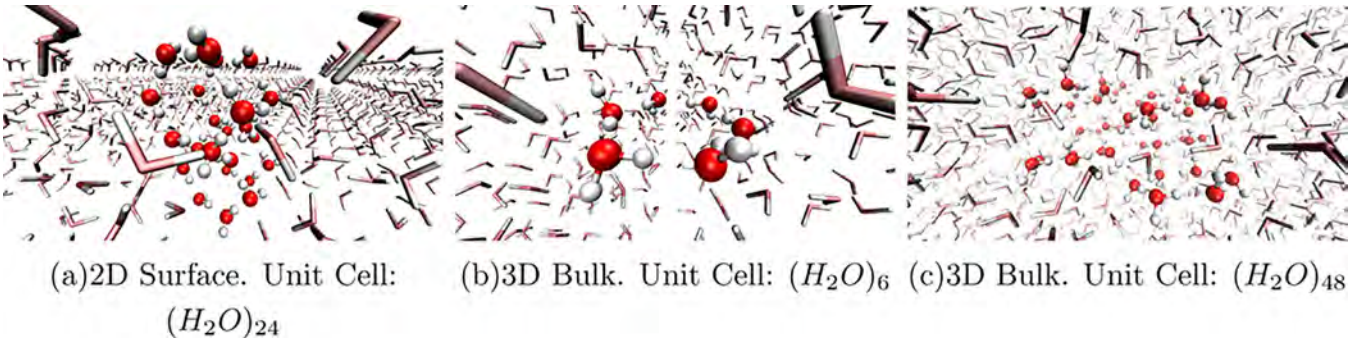


Figure 3. (a−c) Sample structures for the three periodic sets of libraries of pure water in Table 1. Relative energy distributions for each set of structures are displayed in Supporting Information, Figures SI-6 and SI-8.

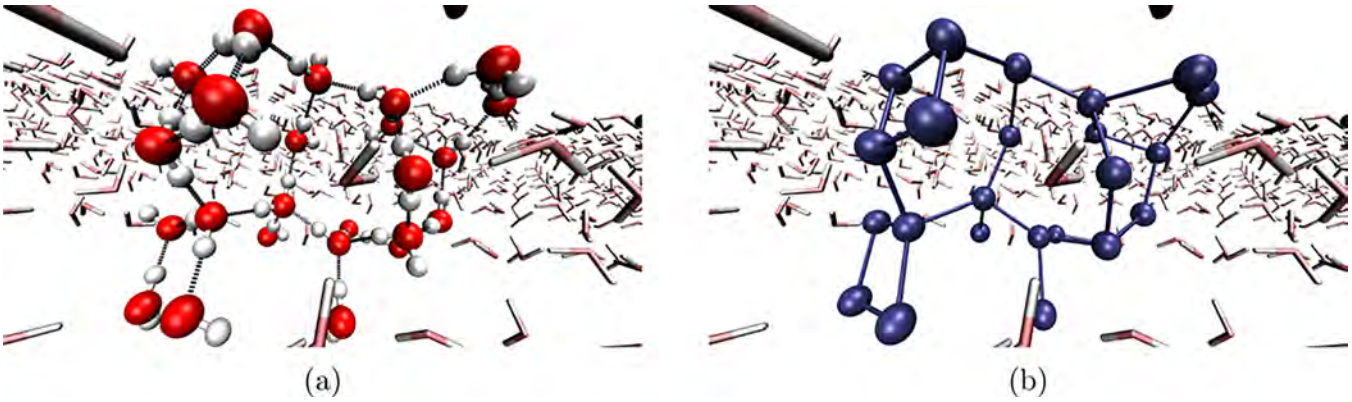


Figure 4. Periodic thin film of water with a unit cell of 24 water molecules is represented in (a), where the hydrogen bonds are depicted with dashed lines. The unit cell region is treated as a graph (b) where the edges are constructed according to the adaptive spatial envelope prescription described in Section 4. This allows the system to be treated by eq 1, treating the periodic character of the system ($E_{\mathrm{PBC}}^{\mathrm{level},0}$) by a standard, affordable method, while exchange−correlation and/or larger basis-set corrections are added by the N-body terms arising from the simplex elements of the graph.

obtained from eq 6, are also preserved in all cases, but eq 17 provides a much more rigorous test than simply gauging the energetic ordering of conformers.

The efficiency of our method is reported as a ratio of the serialized computational run-time of eq 6 and the full-system high-level benchmark serial time. This type of analysis complements the scaling arguments presented previously by us in refs 76 and 79. As noted in ref 98, such an analysis is essential to gauge the utility of such approximate methods. Toward this goal, in this publication, keeping with the terminology previous presented in the literature,[39,88,166−170] the ratio of serialized times mentioned above is referred to in the following subsections as the "speedup". However, as noted above, the C++ module that allows this graph-theoretical computation is MPI-parallelized, given that the computations afforded by eq 6 are trivially parallel.[171,172] For the serial calculation, the full system dominates the CPU times, as seen below and in refs 76 and 78, leading to a choice of parallelization which can reduce the effective wall-time to computational costs of the full-system low-level calculation.

3.2. Construction of Structural Libraries for Condensed-Phase Water Systems. To ascertain the accuracy and efficiency of our methods, we considered a wide-range of structures in the form of bulk water systems conforming to three-dimensional periodic boundary conditions, thin films of water under two-dimensional periodic boundary conditions, and various adsorbate molecules on the surface of water films. The structures utilized to carry out these benchmark studies included both optimized, stable, molecular geometries as well as nonstationary structures obtained during short AIMD trajectories. The Supporting Information document provides an extensive discussion of the bulk and surface structures from the structural library outlined in Tables 1 and 2. Here, we summarize the process of obtaining this library of structures.

Bulk systems with unit cell sizes ranging from 6 to 48 water molecules are considered, and the associated structures are obtained from both optimization calculations through plane-wave basis treatment, as allowed by Quantum ESPRESSO, and from atom-centered Gaussians conforming to Bloch symmetry, as allowed within the Gaussian suite of electronic structure programs. To obtain these structures, our analysis began with the well-studied water hexamer cluster system.[173,174] Lower energy, nonstationary, states were sampled from a set of AIMD trajectories (first row in Table 1, with more details in Supporting Information, Figures SI 3−5). The 15 structures thus generated were used as initial geometries that were replicated and optimized with translation vectors representing the condensed phase. This provided families of surface (two-dimensional periodic boundary conditions, see Figure SI-8 for energy distribution) and bulk (three-dimensional periodic boundary conditions, see Figure SI-6 for energy distribution) structures. The surface structures were optimized with the Gaussian series of electronic structure codes,[73] which uses atom-centered basis functions (the top-block of Table 1), as the procedure did not require periodic calculations in the dimension normal to the surface. The bulk structures were primarily optimized with the Quantum ESPRESSO package,[45] which uses plane-wave basis sets (bottom-block of Table 1). In addition, to make our

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

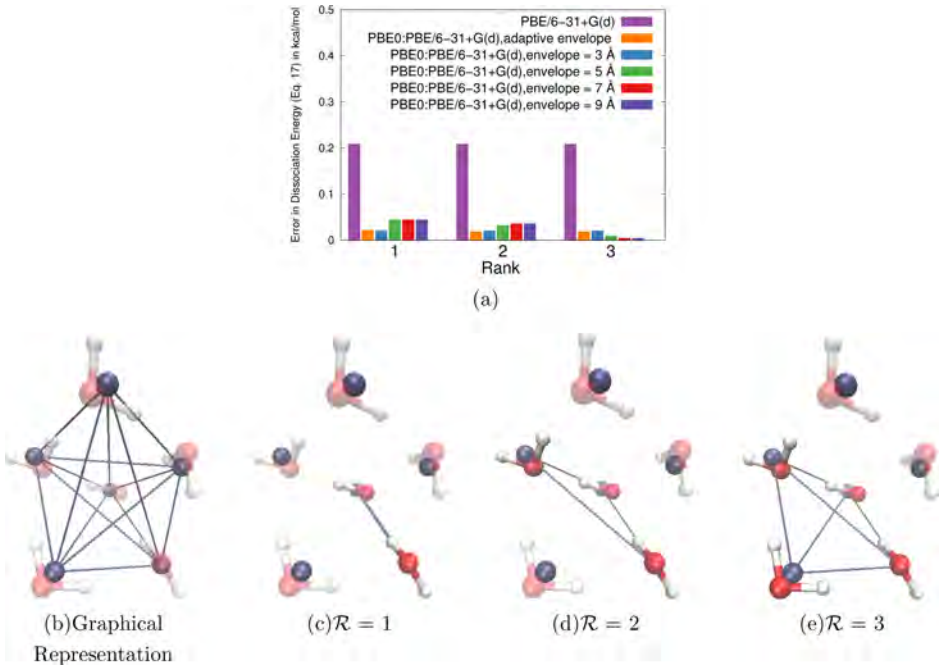


Figure 5. Effect of the two key parameters involved in the construction of the graphical representation in eq 6. These two parameters are the spatial cutoff for edge generation (range of two-body interaction) and the choice of the maximum rank, R (the order of many-body interactions included in eqs 6 and 14). (a) Dependence of error in cluster dissociation energy on both of these parameters is presented here, with more details in the Supporting Information. Here, extrapolation PBE0:PBE/6-31+G(d) → PBE0/6-31+G(d) for the library of $(H_2O)_6$ structures listed in Table 1 is presented and complemented by Figures SI-3 and SI-4. For larger envelopes, the error reduces with increasing rank, but in all cases, the error is well inside the acceptable range. (b) Graphical representation of a 6 water cluster and (c−e) highest rank simplex for a chosen value of $\mathcal{R}$. For example, (e) depicts one specific four-body or rank-3 simplex interaction. These results are complemented by Figure 6 for bulk water systems, and Figure SI-20 shows the respective errors in conformational energy.

benchmarks exhaustive, a 6 water unit cell bulk family was also optimized with atom-centered basis functions from the Gaussian electronic structure code. A few illustrations of these structures are provided in Figure 3.

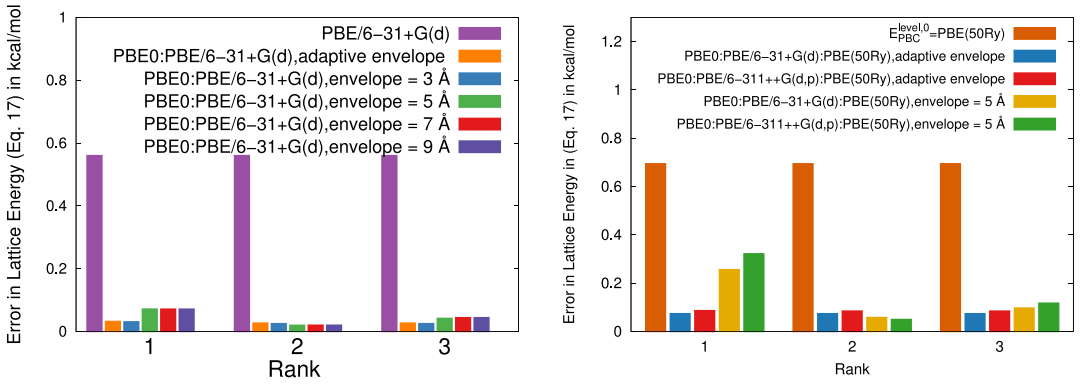
The heterogeneous adsorbate structures were obtained from an elaborate treatment where starting structures were derived from a grid constructed parallel to surface as discussed in Supporting Information Section SI-1. These starting geometries were then optimized to obtain a range of structures used to gauge our methodology. Specifically, the adsorbate molecules (H_2, CH_4, and CH_3OH were considered) were individually positioned on previously optimized water−film interface structures captured in Table 1. The resultant benchmark structures, reflected in Table 2, were obtained through geometry optimization in an atom-centered basis. Next, additional structures were obtained by sampling constant unit cell volume, PBC-AIMD trajectories. For each trajectory, 13 low energy conformations encountered were selected to expand our structural library in Table 2. Additional details can be found in the Supporting Information. Illustrative example structures for the adsorbate-on-water system are provided in Section 5.3.

4. ADAPTIVE MANY-BODY EFFECTS THROUGH GRAPH GENERATION: CHOICE OF SPATIAL ENVELOPE FOR EDGE GENERATION (TWO-BODY INTERACTIONS) AND MAXIMUM RANK OF SIMPLEX, $\mathcal{R}$ (THE ORDER OF MANY-BODY EXPANSION IN EQUATIONS 6 AND 14)

In order to apply eq 6 to the condensed-phase systems discussed above, the unit cell used for the periodic calculations is represented as a graph (see Figures 2 and 4). There are three specific control parameters in eq 6 that influence the construction and evaluation of the graph, the associated many-body approximations in eqs 6 and 14, and the corresponding required set of electronic structure calculations. These include (a) the definition of the node or individual molecular monomeric units that form the basis for the two MBEs in eqs 6 and 14, (b) the spatial envelope that assembles the local family of edges, or two-body interactions, around any given node, and (c) the choice of "$\mathcal{R}$" in eq 6 which dictates the order of many-body interactions to be included.

Each water molecule is designated here as a node within a graph. The nodal coordinates are defined as the algebraic mean of the atomic coordinates within the node. Edges are generated by use of a spatial envelope. Here, two types of spatial envelopes are considered which define the spatial neighborhood used to determine the set of $\{\alpha\}$ for each "r" in eq 6. The first is based upon a fixed distance cutoffs where a given node is connected by an edge to each node within the designated Cartesian distance. This particular feature is similar to that reported in a variety of fragmentation-based methods.[87,91,92,113,114,147,175−178] In addition, the second approach is an adaptive envelope scheme[74] where given the minimum node−node distance (D_i) for a chosen node, edges are formed such that the respective node−node distances are within 10% of D_i. In this way, the graph begins to represent a hydrogen-bonded network for the systems chosen in this publication (as can be seen in Figure 4). However, this network allows us to extract one-body (r = 0, nodal contributions), two-body (r = 1, edge contributions), three-body (r = 2, triangular contributions), and higher-order contributions toward eqs 6 and 14. Furthermore, the specification here is general enough that both bonded and

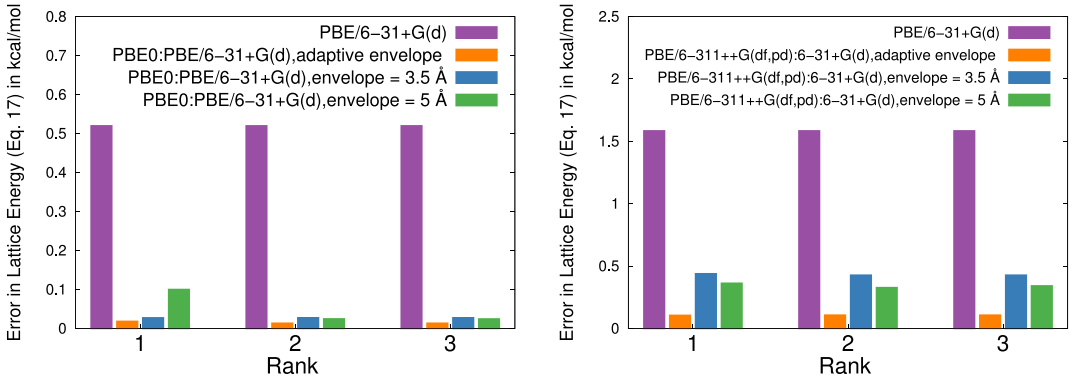
https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812



(a)3D Unit Cell: $(H_2O)_6$: PBE0:PBE/6-31+G(d) → PBE0/6-31+G(d)(Bloch)

(b)3D Unit Cell: $(H_2O)_{48}$: PBE0:PBE/6-31+G(d):PBE(50Ry) → PBE0(50Ry)

Figure 6. Rank and spatial envelope dependence of accuracy of the energy expression in eq 6 is demonstrated using atom-centered basis functions adapted to Bloch symmetry for $E_{PBC}^{level,0}$ (a) and using plane waves for the same (b). As in Figure 5, higher-order many-body interactions provided by larger $\mathcal{R}$ values are not necessary to ensure accuracy for the bulk calculations. Furthermore, an adaptive envelope (adaptive edge generation scheme) is far more effective in capturing these interactions. This study is expanded upon in the Supporting Information for a variety of unit cell sizes and functionals (Figures SI-14 and SI-15), and Figure SI-20 shows the respective errors in conformational energy as a counterpart to (a).



(a)2D Film Unit Cell: $(H_2O)_{24}$: PBE0:PBE/6-31+G(d) → PBE0/6-31+G(d)(Bloch)

(b)2D Film Unit Cell: $(H_2O)_{24}$: PBE/6-311++G(df,pd):6-31+G(d) → PBE/6-311++G(df,pd)(Bloch)

Figure 7. Similar to Figure 6 but for the inhomogeneous surfaces. (a) Effect of rank and spatial envelope on functional extrapolation, whereas (b) effect of rank and spatial envelope on basis-set extrapolation for water surfaces. Panel (a) is complemented by Figure SI-16 where an exhaustive set of functionals is considered and Figure SI-21 which also provides the conformational energy error for (a).

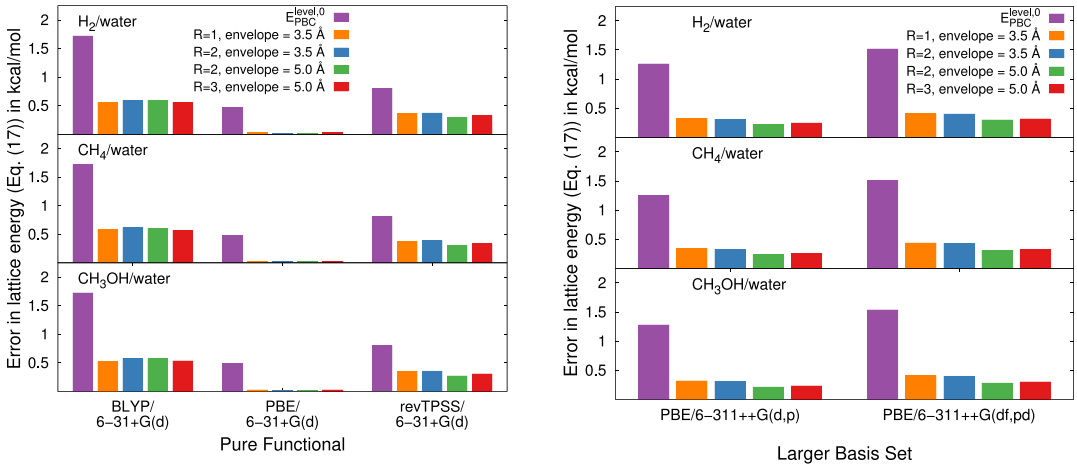
nonbonded interactions can be considered easily using the protocols described above and have been demonstrated as part of AIMD simulations in refs 74 and 78.

The nodes and edges together define the graph. Thus, in summary, a local interaction neighborhood is created using spatial envelopes, either adaptive or fixed cutoff, to compute E_χ in eq 6 which captures the many-body interactions in the system up to order $\mathcal{R}$ and implemented within eq 6. This neighborhood graph creation algorithm is implemented using Python functions called from a C++ module which allows the automated formation of these graphs and the associated molecular fragments up to order $\mathcal{R}$ in eq 6 for arbitrary (bonded and nonbonded[74,76−78]) systems (see eqs 12 and 13 regarding connections to many-body expansions).

4.1. Rank ($\mathcal{R}$) Dependence of Accuracy of Equation 6. In order to gauge the quality of eq 6, $\mathcal{R}$, the maximum simplex rank, and the spatial envelope, for edge creation, need to be chosen for the subsequent condensed-phase studies. In this section, we consider the use of eq 6 by varying both $\mathcal{R}$ and spatial envelope sizes. This is first done on gas-phase clusters of six water molecules and then on condensed-phase systems that include both homogeneous bulk and heterogeneous adsorbate interaction systems. The effect on functional extrapolation and basis extrapolation is considered. The gas-phase systems help focus on the local many-body expansion without condensed-phase effects. The structures used are described in Section 3.2 and in greater detail in Supporting Information Section SI-1.

In Figure 5, we present errors in cluster dissociation energy as a function of both $\mathcal{R}$ and spatial envelope sizes for the water clusters. The dissociation energy is the gas-phase equivalent of eq 16. It is found that the graphical representation of these water structures only required $\mathcal{R} = 1$ contributions (Figure 5c) along with the adaptive envelope protocol. Although rank-3 objects (Figure 5e) for these clusters do demonstrate even greater

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812



(a)Adsorbate Unit Cell: $(H_2O)_{24}$. Target: PBE0/6-31+G(d)(Bloch)

(b)Adsorbate Unit Cell: $(H_2O)_{24}$. Extrapolating from PBE/6-31+G(d)

Figure 8. (a) Similar to Figure 7a but for adsorbate systems. The horizontal axis represents the level, 0, whereas the target level, 1 is presented in the subfigure caption. Panel (b) is analogous to Figure 7b but again for adsorbate systems. In this case, the horizontal axis refers to the target level, 1 study, whereas level, 0 is presented in the subfigure caption. These are further elaborated upon in Figure SI-19.

accuracy when enabled by larger spatial envelopes, the accuracy afforded by the rank-1 objects was found to be sufficient.

Next, Figures 6−8 portray the accuracy of eq 6 for a variety of $\mathcal{R}$ values and spatial envelope choices, but for the bulk water systems, surfaces, and adsorption problems. As was seen in Figure 5 for the water cluster case, accurate lattice energies can be achieved with $\mathcal{R} = 1$ and the use of the adaptive envelope protocol. A similar analysis is conducted in the Supporting Information section. Specifically, for a range of bulk water systems, we study the accuracy of functional extrapolation in Figures SI-14 and SI-15 (which complements the results in Figure 6b). In all cases, a variety of $\mathcal{R}$ values and spatial envelope choices are considered. It is found that the effect of higher-order many-body terms is largely negligible because of the ONIOM-like character of the approach; the higher-order contributions are already included to some extent at the base-level determined by $E_{PBC}^{level,0}$. To further expound upon this study, we also consider the influence of rank and spatial envelope on inhomogenous water surfaces and adsorbate−water interfaces.

A few comments are in order with respect to the adaptive edge generation. For all systems treated in Figures 6−8, the fixed hard edge cutoffs of 5 Å and longer show a larger error as compared to the adaptive edge generation scheme[74] when only edges ($\mathcal{R} = 1$) are considered. With the inclusion of $\mathcal{R} = 2$ effects (three-body terms), these longer distance interaction are better described. This is, however, not found to be necessary within the adaptive edge generation scheme.[74] The improvement in the performance from the fixed edge study is partly due to the fact that the longer edge distances include nondirect hydrogen bonding water pairs without the intermediate mutually hydrogen-bonded neighbor. With the inclusion of the three-body terms, these cooperative effects are properly accounted for through the mutual hydrogen-bonded neighbor. Aside from the above consideration, the increase of $\mathcal{R}$ does not significantly improve the accuracy even with the increase of the unit cell size and certainly has a minimal effect within the adaptive scheme. The results provided in Figures 7 and 8 are further elaborated in Supporting Information, Figures SI-16 and SI-19. At the end, $\mathcal{R} = 1$ with adaptive edge generation appears to be a good stable choice that was used for the remaining portion of this paper.

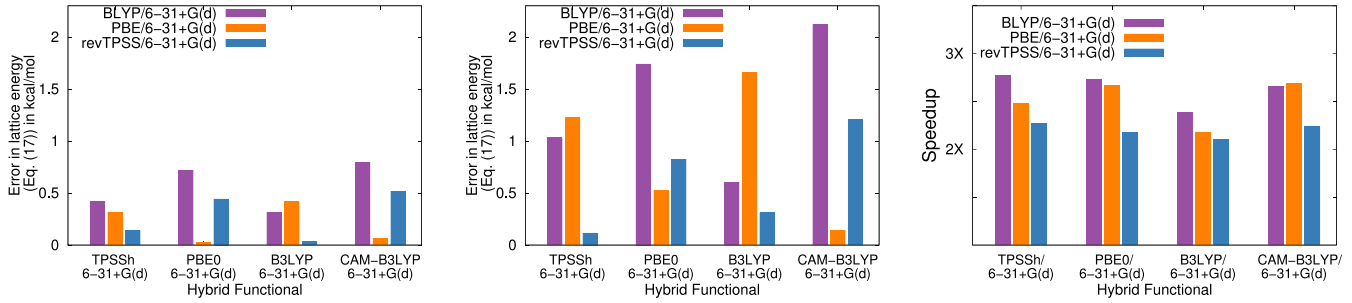
The results here are also consistent with the discussions in ref 94 for finite cluster systems. From these discussions, it follows that methods that retain a strong ONIOM-type basis of treating the full-system at a lower-level of electronic structure may have similar characteristics. Such methods include the multicentered QM/QM formalism,[81,179] the molecular tailoring approach,[83] ONIOM-XO,[82] the molecules-in-molecules methodology,[84] and HMBI.[86,129] These aspects have also been subsequently noted in ref 98. The use of standard MBE methods with electronic embedding (i.e., unlike within the ONIOM-like description here), by contrast, often require higher-order terms to converge to the correct energy.[18,71,72,114,180−182] This requirement of higher-order terms in noncomposite methods,[94,95,98] many-body expansions with[90,91,99,106,121,180] and without[68−72,102] electronic embedding, can be further exasperated by basis-set superposition error from the fragments.[183]

The larger rank terms may still be required when (a) the chosen full-system low-level treatment does not sufficiently capture many-body effects in a given system or (b) many-body effects involve true long-range correlation. Because of the findings above, the production calculations presented in this paper (in the next section) use $\mathcal{R} = 1$ representations, and the edges (or two-body terms) are generated using an adaptive envelope for the pure water systems. In practice, this choice implies that eq 7 is employed which includes two-body corrections to exchange, correlation, and large basis effects based upon an adaptive distance protocol, and the required higher-order many-body effects are captured within $E_{PBC}^{level,0}$.

5. EFFICIENT EVALUATION OF LARGE BASIS-SET APPROXIMATIONS AND RUNG-4 EXCHANGE−CORRELATION FUNCTIONALS FOR CONDENSED-PHASE SYSTEMS USING EQUATION 6

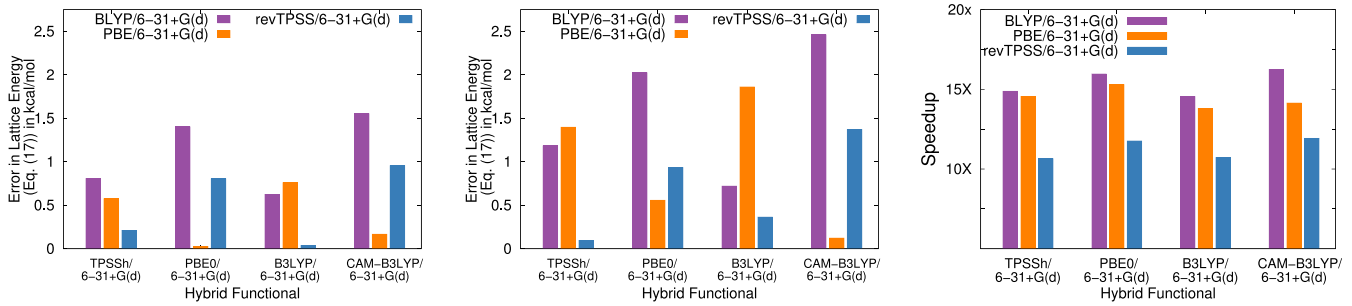
Hybrid density functional methods include a fraction of Hartree−Fock exchange and are computationally expensive for periodic calculations because of their spatially nonlocal nature.[11,41−43,184−186] Furthermore, for plane-wave methods

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812



(a) $E^{graph-theor.}_{PBC,R=1}$ versus $E^{level,1}_{PBC}$ (b) $E^{level,0}_{PBC}$ versus $E^{level,1}_{PBC}$ (c) Speedup

Figure 9. Here, the accuracy of the energy expression in eq 7 is evaluated using eq 17 for the films of water (two-dimensional periodic boundary conditions) with a unit cell of size 24 water molecules (Figure 2). In (a), the target (rung-4) functionals are presented along the horizontal axis, and levels used to compute $E^{level,0}_{PBC}$ (which dominates the computational effort) are shown on the legend. Clearly, the lattice energy errors are fractions of kcal/mol. The baseline comparisons (b) of the chosen lower levels of theory with the high levels of theory show lattice energy error of the order of 1−2 kcals/mol, with some exceptions discussed in the text. Very similar results were recovered for films with unit cell sizes containing 12 water molecules, as shown in Figure SI-17. Similar extrapolations using larger $\mathcal{R}$ and spatial envelopes are shown in Figure SI-16. A similar plot which shows the error in conformational energy is presented in Figure SI-21. The speedup from using eq 7 is presented in (c). Here, the speedup is defined as the ratio of the serialized computation time for calculations performed at the higher level of theory on the periodic systems and those obtained using eq 7. A clear speedup factor of 2−3 is noted for these homogeneous water surfaces.



(a) $E^{graph-theor.}_{PBC,R=1}$ versus $E^{level,1}_{PBC}$ (b) $E^{level,0}_{PBC}$ versus $E^{level,1}_{PBC}$ (c) Speedup

Figure 10. Similar to Figure 9 but for bulk water. (a) Each of the target hybrid functionals is extrapolated to within a fraction of a kcal/mol with an appropriately chosen semilocal DFT method. The effect of this extrapolation on conformational energy is shown in Figure SI-22. The efficiency of these extrapolations leads to roughly a factor of 15 speedup, as seen in Figure 10c (even greater speedups are obtained when extrapolations are done using plane-wave basis sets in Figure 11b). Baseline comparisons with level, 0 PBC calculations are presented in (b).

exact exchange requires quadratically greater Fourier transforms than for the semilocal approaches.[19,187,188] The pure, semilocal DFT functionals do not suffer from this computational challenge but are considered to be less accurate in evaluating hydrogen-bonded and other noncovalent interactions.[13,33,38,189,190] Pure functionals also suffer from self-interaction of electrons,[189,191−193] which is significantly reduced by the fractional contribution of the exact exchange incorporated into hybrid methods. Equation 6 partially alleviates this by obtaining approximations to higher rung functionals at a reduced cost. This approach is employed for the study of thin films (Figure 9), bulk water (Figures 10 and 11), and the surface adsorption systems (Figure 15).

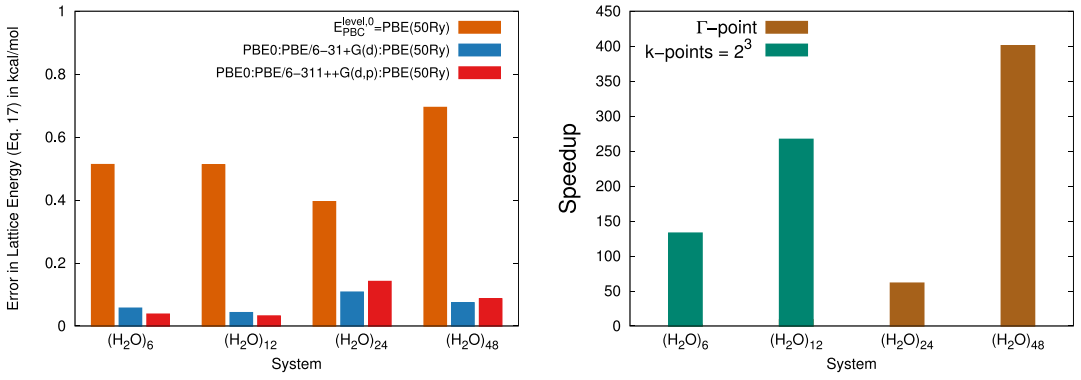
Our methodology shares similar objectives as range-specified "screened" exchange functionals.[41−43,184,186] However, here, the extent of locality captured by the Hartree−Fock exchange may be tailored adaptively and made spatially dependent using graph theory. Specifically, this may be done through the inclusion of (a) a position, or node, dependent spatial envelope that defines the edges, and hence the graph, as outlined in Section 4, and (b) the value of maximum simplex rank, $\mathcal{R}$, in eq 6. As noted in Section 4.1, for the systems studied here, it was found that these higher-order terms are not necessary to capture the lattice energies for condensed-phase water systems. In the studies presented below, we truncated eq 6 at $\mathcal{R} = 1$ and used the adaptive spatial envelope as explained in Section 4.1.

This section is organized as follows: Section 5.1 presents an analysis of errors and efficiency for condensed-phase systems, with more details in the Supporting Information. Basis-set extrapolation is considered in Section 5.2 and surfaces with organic solutes are considered in Section 5.3. In all cases, the error in lattice energy is analyzed using eq 17.

5.1. Efficient and Accurate Hybrid (Rung-4) DFT for the Condensed Phase. In Section 5.1.1, we begin with a combinatorial approach by extrapolating from three pure functions to four hybrid functionals using atom-centered basis functions with Bloch symmetry. Next, in Section 5.1.2, using an advantageous extrapolation pair (PBE0:PBE) from the study in Section 5.1.1, a further set of calculations are performed with increasing unit cell sizes of bulk water structures, treated with plane-wave basis sets.

5.1.1. Combinatorial Benchmarks Using Atom-Centered Basis Functions Adapted to Bloch Symmetry. In light of eq 1, our method can be seen as a graphical, or many-body, correction to the chosen low-level treatment in order to approach the accuracy of the target high-level treatment. As such, this method

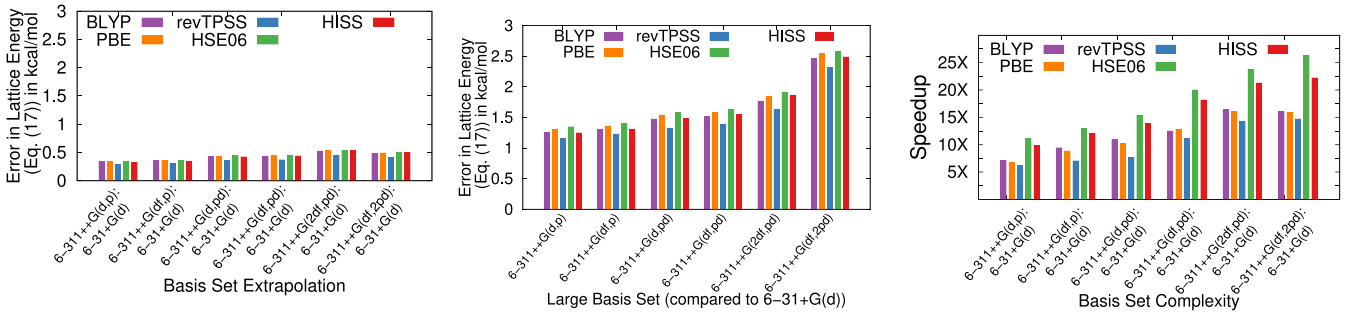
https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812



(a)PBE0:PBE/6-31+G(d):PBE(50Ry) → PBE0(50Ry) and PBE(50Ry) versus PBE0(50Ry)

(b)Speedup

Figure 11. (a) Similar results to that in Figure 10, but now the full-system low level is treated using a plane-wave basis with a 50 Rydberg kinetic energy cutoff and the Quantum ESPRESSO electronic structure package.[45] As specified in Table 1, the 6 and 12 water unit cells used a k-mesh of 2^3, whereas 24 and 48 used the Γ-point [as can be seen in Table 1 and (b)]. This is true for all bulk calculations including the full-system low level. The expanded treatment of each of these systems, including the effect of choice of maximum rank of simplexes and the envelope size, is provided in Figures SI-14 and SI-15 and the conformational errors are provided in Figure SI-22. (b) Extrapolation PBE0:PBE/6-31+G(d):PBE(50 Ry) → PBE0(50 Ry) offers significant speedup with increasing unit cell size. For the 48 water unit cell systems, these calculations were able to reduce the computational times from about 5 days to 15 min serial CPU time (or 3.5 min with MPI parallelism). The study here provides a robust and efficient option to add hybrid corrections to plane-wave calculations.



(a)$E_{PBC,R=1}^{graph-theor.}$ (Eq. (7)) versus $E_{PBC}^{level,1}$

(b)$E_{PBC}^{level,0}$ versus $E_{PBC}^{level,1}$

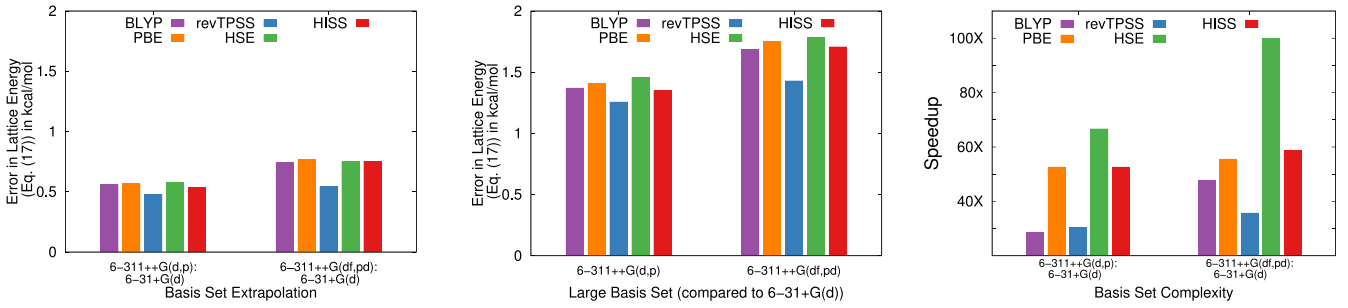
(c)Speedup

Figure 12. Here, the accuracy of the energy expression in eq 7 is evaluated using eq 17 for large basis calculations with two-dimensional periodic boundary conditions. The unit cell size of 24 water molecules. The target basis and the basis sets used to compute $E_{PBC}^{level,0}$ are presented along the horizontal axis, while the functionals are shown on the legend. Clearly, the lattice energy errors from the graph-theoretic formalism (a) are fractions of kcal/mol. The baseline comparisons (b) for the chosen smaller basis set, 6-31+G(d), with the larger basis sets for the film of water molecules shows a mean absolute error on the order of 1−3 kcal/mol per water (similar results were found for the 12-water unit cell, as seen in Figure SI-18). Similar extrapolations using larger $\mathcal{R}$ and spatial envelopes are shown previously in Figure 7b. The choice of $\mathcal{R}$ is shown to have minimal impact, and the adaptive cutoff offers the best accuracy for basis-set extrapolation, consistent with the earlier discussion. The analogous plots for the conformational energy errors are provided in Figure SI-23. The speedup from using eq 7 increases from a factor of 6 to a factor of 20 as the target large basis set is increased and is reported in (c).

provides accuracy according the target (hybrid DFT) level of treatment, at computational cost commensurate with the chosen lower (pure functional) level of theory for condensed-phase systems. Thus, to provide a baseline error analysis, we also present the mean absolute lattice energy errors between the low (level, 0) and high (level, 1) levels of theory in Figures 9b, 10b, and 11a. The chosen low levels of theory (BLYP,[62] PBE,[63] and revTPSS[64]) varied in how well they recovered the higher level (TPSSh,[60] PBE0,[61] B3LYP,[58] and CAM-B3LYP[59]), and our next goal is to check if this is improved upon by using the methods presented above. A consistent atom-centered basis set was used, 6-31+G(d), for the calculations in Figures 9 and 10, meaning only the density functionals are modified while the basis-set treatment remains constant.

Equation 17 is employed to compare the hybrid DFT lattice energy results to that of both the lattice energies from the low-level baseline semilocal functionals and from the use of eq 7 (in light of Section 5.1.1) to extrapolate from those low-level functionals. When we consult the baseline results in Figures 9b and 10b, we note that TPSSh lattice results are already well captured by the modified version[64] of its parent 3rd rung functional TPSS.[32] Furthermore, the lattice energies from CAM-B3LYP are surprisingly well captured by the 2nd rung PBE functional. Meanwhile, PBE0 and B3LYP are less well represented by these lower rung functionals, but PBE0 appears

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812



(a) $E_{PBC,R=1}^{graph-theor.}$ versus $E_{PBC}^{level,1}$ (b) $E_{PBC}^{level,0}$ versus $E_{PBC}^{level,1}$ (c) 3D Bulk. Unit Cell: $(H_2O)_6$

Figure 13. Similar to Figure 12 but for bulk water. The use of eq 7, in (a), shows a significant gain in accuracy over the baseline studies shown in (b). The conformational energy errors for these extrapolations are provided in Figure SI-24. Although the errors remain under a kcal/mol upon use of eq 7, the computational efficiency is substantially greater and shows a 30−50 speedup as seen in (c). The speedup is computed with respect to the same calculations performed at the higher level of theory on the systems with a unit cell size of 24 water molecules.

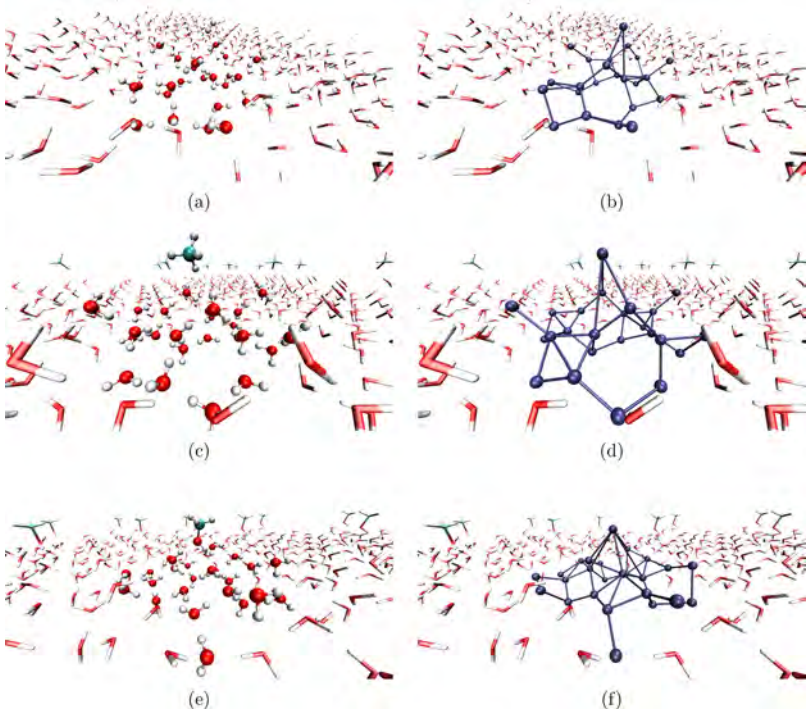


Figure 14. Adsorbates considered in this study are H_2 (a,b), CH_4 (c,d), and CH_3OH (e,f). The graphical representations, displayed in (b,d,f), are localized to the unit cell in this study (although this is not a requirement) and used for the evaluation of eq 7. For the case of methanol (e,f), it is clearly seen that the graphical decomposition considers several nonbonded interactions.

to be nicely approximated by its parent functional PBE, whereas B3LYP is moderately matched by revTPSS.

The next question pertains to the extent of the improvement in accuracy when using eq 7. Overall, the use of this edge-based (two-body) extrapolation demonstrates significant improvement in obtaining the target hybrid functional lattice energies, with the notable exceptions of TPSSh:revTPSS/6-31+G(d) → TPSSh/6-31+G(d)(Bloch) and CAM-B3LYP:PBE/6-31+G(d) → CAM-B3LYP/6-31+G(d)(Bloch), as the low-level calculations reproduce the calculated lattice energies reasonably well. The best extrapolations tested here were PBE0:PBE/6-31+G(d) → PBE0/6-31+G(d)(Bloch) and B3LYP:revTPSS/6-31+G(d) → B3LYP/6-31+G(d)(Bloch), where the low level already demonstrated a fair degree of accuracy in replicating the target energies but with the edge corrections from eq 7 demonstrates that a remarkable recovery of the lattice energy is achieved in both bulk water (Figure 10a) and surface (Figure 9a) calculations at much reduced costs. This reduction in cost is quantified in Figures 9c and 10c where the speedup is the ratio of CPU time cost of the target functional over the computational time cost of the use of eq 7. For the water film systems, the reduction in computational cost is by more than a third, whereas for the three-dimensional bulk system, we see a roughly 15 times reduction in computational times. (In the next paragraph, we find that the speedup is much greater when plane-wave basis functions are used.) This manner of reporting does not consider the run-time improvement gained through parallel treatment of the individual component calculations.

5.1.2. Plane-Wave Basis Treatment of Larger Unit Cells. Because the extrapolation PBE0:PBE/6-31+G(d) → PBE0/6-31+G(d)(Bloch) proved to be advantageous in Figures 9 and 10, this combination of functionals was again considered for a

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

series of larger bulk systems in Figure 11. We consider a similar scheme as before but replace $E_{PBC}^{level,0}$ in eq 7 with the corresponding energy obtained from plane-wave basis treatment from Quantum ESPRESSO.[45] The terms $E^{level,0}(\alpha,r)$ and $E^{level,1}(\alpha,r)$ remain in an atom-centered Gaussian basis using the same DFT functional considered in the Quantum ESPRESSO calculation. In Figure 11, we use a plane-wave basis with a kinetic energy cutoff of 50 Ry using the projector-augmented wave (PAW) representation[46] with the Kresse−Joubert style in both the periodic high (target) and low levels (level, 0) of treatment. The fragments, as noted at the top of Section 5, used the atom-centered basis functions, 6-31+G(d) and 6-311++G(d,p). The *k*-mesh chosen for both the pure and hybrid DFT periodic treatments are indicated in Table 1. Note that when the atom-centered basis calculations are performed the choice of *k*-mesh differs between pure and hybrid functionals. However, for the plane-wave basis, the pure and hybrid functionals both have the same number of *k*-points, and the chosen *k*-mesh reduces to the Γ-point for the larger bulk systems (Figure 11). The reason for this difference can be seen from Figure 1 where the computational expense for *k*-sampling in hybrid functional calculations is shown to be prohibitive, especially for plane waves.[39,44,194]

As was seen in the extrapolation to the atom-centered periodic treatment of bulk water at PBE0 in Figure 10, significant agreement with hybrid plane-wave lattice energies was also achieved as compared to the standard pure DFT treatment (orange bars in Figure 11a) for the plane-wave case. This extrapolation proved to be accurate for all unit cell sizes considered (6 water molecules to 48 water molecules) and is able to reduce a PBE0 calculation from approximately 5 days down to a few minutes bottlenecked at PBE(50 Ry) (Figure 11b). The effect of rank and envelope sizes has been discussed in Section 4.1 and is expounded upon in the Supporting Information. As noted, these effects are found to be minimal.

In summary, we are able to achieve sub-kcal/mol accuracy for higher levels of density functional methods at a fraction of computational cost. This approach is shown to work well with both bulk and condensed-phase surfaces and with both plane-wave and Bloch symmetrized atom-centered bases.

5.2. Efficient and Accurate Basis-Set Extrapolation with Periodic Systems. When using nonorthogonal atom-centered Gaussian basis sets in periodic simulations, the electronic basis functions are transformed into orthogonal and translationally invariant Bloch functions.[195] These are also often referred to as "crystalline orbitals".[11] However, when diffuse and polarized atom-centered basis functions are included, significant density may be found either (a) outside of the periodic cell or (b) in regions with limited atomic density, where the latter leads to instabilities in the SCF procedure.[11,165] Furthermore, from using Ewald summation[57] and from fast multipole-type approximations,[54,55] the long-range portion of the Coulomb repulsion term is computed in reciprocal space (*k*-space) for improved efficiency; the associated integration over *k*-cells contributes in a significant manner to the cost of expanding the chosen (diffused and polarized) basis-set size. However, larger basis sets are often needed to represent weak interactions[78,196] that are present in catalytic problems.[6,8,197] To achieve stable and efficient basis-set treatments of condensed-phase reactive systems, we employ eq 6 as done in ref 78. By use of the graph-theoretic treatment discussed above, we are able to obtain large basis treatment of local interactions. Local is defined by the graphical decomposition perspective in eq 6 and is quantified by (a) the size of $\mathcal{R}$ and (b) the density of local nodal connectivity through edges. Facilitated by the discussion in Section 4.1, here we consider two-body-based local basis-set space expansion (see eq 14) which is represented as edges in the graphical representation of the system. This is done in conjunction with treating the condensed-phase portion, $E_{PBC}^{level,0}$ in eq 6, using less computationally intensive atom-centered basis sets to reduce cost and increase stability with periodic *k*-cell integration (see Tables 1 and 2). Here, we provide a numerical tool to compute properties accurate to highly diffuse Pople-type basis sets (such as 6-311++G(df,2dp)) at costs commensurate with minimal Pople-type basis sets such as 6-31+G(d).[198] Similar basis-set extrapolations, using the same low-level basis set, were shown to perform well in ref 78 for molecules in the gas phase. As we demonstrate below, this reduces the computational costs and is hence a key step in pushing the realm of possible accurate condensed-phase simulations.

Graph theory-based basis-set computations were performed on the thin films for multiple DFT methods, and these are shown in Figure 12. The selected DFT methods are the three semilocal functionals considered in Section 5.1.1 and two screened hybrid GGA functionals, HISS and HSE,[41−43] which employ a distance-based screening of the exact exchange for reduced computational cost in the condensed phase. Through computational effort commensurate with 6-31+G(d), we are able to obtain accuracy comparable to multiple diffuse basis sets with accuracy in lattice energy in the sub-kcal/mol range (see Figure 12). The lattice energy errors remain consistent across irrespective of choice of DFT functional which is an attestation to the robustness of this approach and shows that the method is transferable across DFT methods. Furthermore, as noted in Figure 12c, the graph-theoretic procedure provides a 6-25 speedup as compared to PBC calculations using large basis sets on water surfaces. Bulk water systems, shown in Figure 13, have also been treated in a similar manner. Here, again we observe sub-kcal/mol accuracy, with a computational speedup of the order of 25−50 as compared to the large basis full-system calculations. In summary, we have demonstrated the accuracy and significant efficiency of our graph-theoretic procedure to capture many-body interactions on surfaces and condensed phase. This is the case for both accurate density functional computations and for large basis-set calculations.

5.3. Error Analysis for the Study of Organic Impurities on the Surface of Water. Next, we consider the adsorption of H_2, methane, and methanol, on a film of water. The goal here is to gauge the effectiveness of the scheme tested above for the study of heterogeneous systems that may be of significance in chemical catalysis.[5,6,8,199,200] The structures considered here are obtained from both geometry optimization calculations and from sampling AIMD trajectories of systems where the unit cell is chosen to contain 24 water molecules interacting with a single adsorbate molecule, as described in Section 3.2 and the Supporting Information. The graphical representation of the system (Figure 14) is constructed along a similar vein as that described in Section 4, but in this case, the adsorbed impurity is treated as a single additional node that forms edges based upon a spatial envelope with a cutoff of 5.5 Å. The graphical representation capturing water−water interactions was maintained along similar lines as those described in Section 4. Figure 14 provides an example set of graphical representations. Because of the presence of the impurity on the surface, one may suspect that graphical faces (representing three-body interactions) involving two water molecules and a solute molecule are

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

significant. However, the effect of $\mathcal{R} = 2$ and $\mathcal{R} = 3$ terms are shown to be relatively minimal. See Figure 7 and associated figures in the Supporting Information.

The set of DFT extrapolations tested in the previous subsection are also considered here for the solute−water−surface interaction, and the results are shown in Figure 15. The

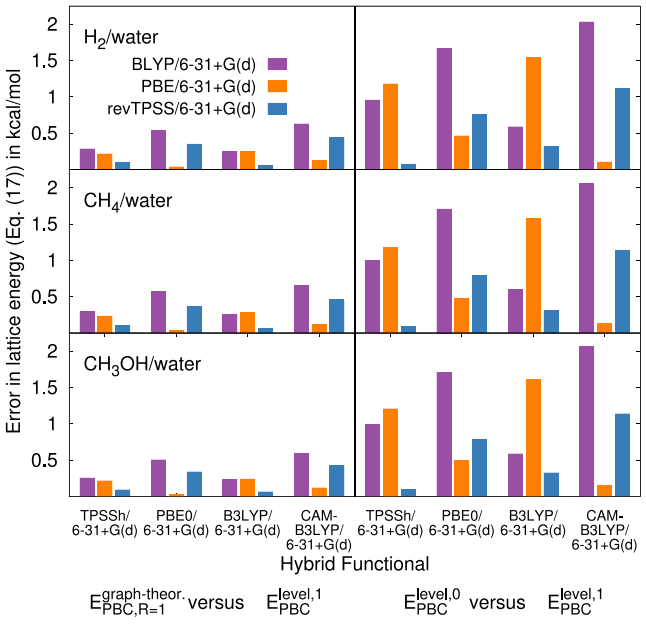


Figure 15. Here, the accuracy of the energy expression of eq 7 is evaluated using eq 17 for a film of water with one solute molecule on its surface. The unit cell consists of 24 water molecules with one solute molecule (H_2, CH_4, or CH_3OH) as in Figure 14. The target (rung-4) functionals are presented along the horizontal axis, whereas the solute molecule is presented above the histogram plots on the left. The use of eq 7 provides accurate lattice energies as seen from the figures on the left. These improvements are clarified by providing baseline comparisons with the respective lower level functionals on the right side. Furthermore, Figures 7 and SI-19 provide an expanded study of convergence of eq 6 as a function of $\mathcal{R}$, and the $\mathcal{R} = 1$ effects captured in eq 7 appear to appropriately include the effects presented here.

trends here are similar to those for the pure water film and bulk systems in Figures 9 and 10. As before, the most beneficial extrapolations are PBE0:PBE/6-31+G(d) → PBE0/6-31+G(d)(Bloch) and B3LYP:revTPSS/6-31+G(d) → B3LYP/6-31+G(d)(Bloch). These calculations show an approximate 2−3 speedup (as was seen in Figure 9c for the homogeneous surfaces) as compared to the full system with a slight increase in efficiency with the size of the adsorbed solute, as seen in Figure 16.

The basis-set extrapolations (Figure 17) show lattice energy errors of less than 0.5 kcal/mol for all five functionals, which is a

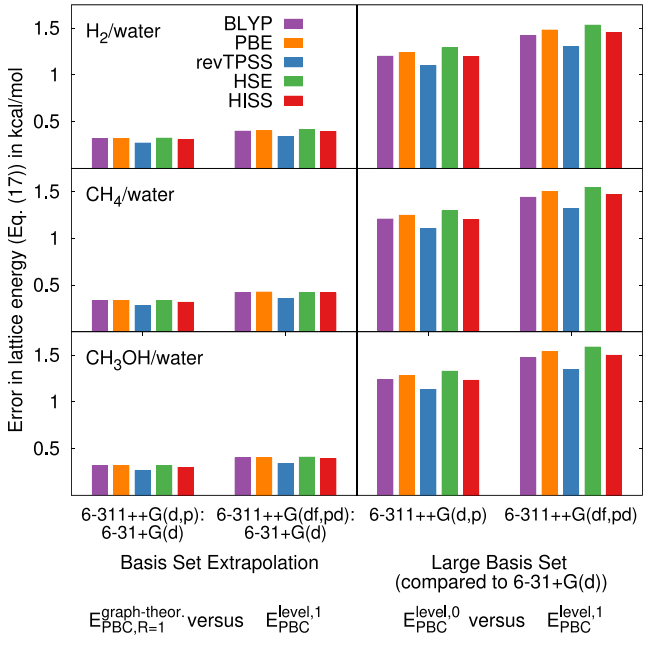


Figure 17. Similar to Figure 15 but for basis-set extrapolation. As before, we choose the basis set 6-31+G(d) as the lower level and these are used within the graph-theoretic formalism in eq 7 to provide results for higher level basis sets presented along the horizontal axis. The solute molecule for each set of systems is presented above the histograms on the left. The graph-theoretic basis-set extrapolations (left) consistently show accuracy to within 0.5 kcal/mol in lattice energy, while the smaller basis baseline (right) shows nearly 4 times the error. The speedup is large as shown in Figure 18. In Section 4.1, Figure 8b, we provided an expanded study of convergence of eq 6 as a function of R, and the $R = 1$ effects captured in eq 7 appear to appropriately represent the effects presented here.

significant improvement when compared with the baseline results shown in Figure 17. As noted above, these results hold across a variety of functionals. A significant cost reduction is seen in Figure 18 with increasing relative efficiency with target basis-set size and through consideration of screened hybrid GGA functionals HISS and HSE.

In summary, we have demonstrated that the methods we have developed are equally applicable for pure water condensed-phase system and to heterogeneous systems such as adsorbates

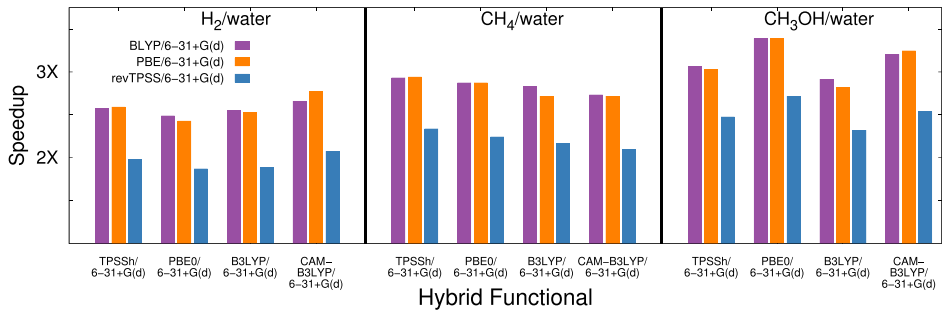


Figure 16. Speedup through the use of eq 7. The speedup is computed with respect to the same calculations performed at the higher level of theory on the water films with solutes. The semilocal DFT to hybrid DFT extrapolations demonstrated a speedup of about a factor of 2−3.

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

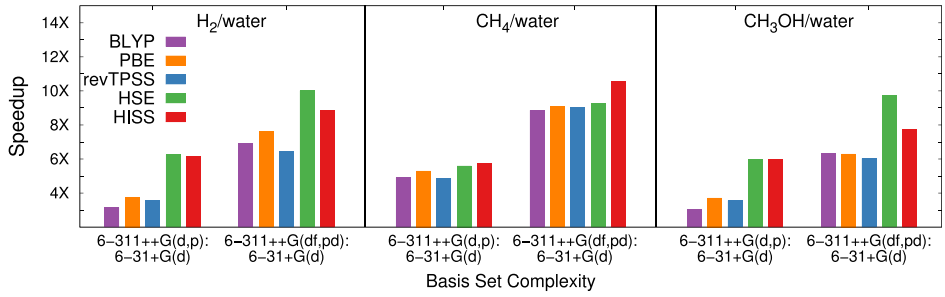


Figure 18. Speedup through the use of eq 7. The speedup is computed with respect to the same calculations performed at the higher level of theory on the water films with solutes. Basis-set extrapolation demonstrates a speedup of roughly a factor of 5−10 as the size of the solute studied increased.

on the surface of water. This application also demonstrates an increase in relative efficiency over the target calculations, allowing future studies on complex heterogeneous systems with greater efficiency and stability.

6. CONCLUSIONS

We discuss the implementation of our graph-theoretic extrapolation procedure for accurate condensed-phase electronic structure studies in homogeneous as well heterogeneous systems. This was achieved by partitioning a chosen cluster within the condensed-phase system into vertices or graphical nodes, that are then connected through edges to form a geometric graph representation or interaction network. This is then used to construct an expanding series of local many-body interactions up to arbitrary orders. Such an approximation is then embedded within the ONIOM extrapolation scheme to arrive at an overall energy expression that provides a conduit for efficient and accurate condensed-phase treatment. Specifically, the graph-theoretic partitioning scheme provides a perturbative correction to an affordable periodic electronic structure method leading to improved accuracy and efficiency in functional and basis-set extrapolation. The accuracy of this approach was demonstrated for both plane-wave and atom-centered Gaussian basis functions adapted to Bloch symmetry.

Specifically, this paper demonstrates three significant observations. First, semilocal DFT results may be used in conjunction with the above-mentioned approach to yield hybrid DFT accuracy in the condensed phase, at "pure" DFT cost for both plane-wave and atom-centered Gaussian basis functions adapted to Bloch symmetry. Second, modest sized atom-centered basis-set treatment may also be used with this approach to achieve large basis-set accuracy, at much reduced cost. Third, low-rank approximations emanating from the aforementioned graph-theoretic many-body theory are sufficient to describe condensed matter systems studied here to obtain results in good agreement with computationally prohibitive condensed-phase hybrid DFT and large basis calculations for bulk systems as well as adsorption surfaces, that is, homogeneous as well as heterogeneous systems. The third observation has been shown to be the case regardless of basis (plane waves and atom-centered Gaussians) and independent of software package; indeed all calculations performed here use both Quantum ESPRESSO as well as Gaussian for a single-energy evaluation for the specific condensed matter system. In all cases sub-kcal/mol lattice energy errors are shown to be achievable at significant reduction in computational cost. The accuracy and efficiency of this method was the most pronounced when considering bulk water (e.g., with $(H_2O)_{48}$ unit cell) within a plane-wave treatment of periodic boundary conditions, achieving a computational cost reduction from a few days down to a couple minutes.

The systems considered as part of our computational benchmarks include bulk water systems with unit cell sizes of up to 48 water molecules, and water surfaces, films, and those involving the interfacial interaction of these surfaces with organic adsorbates. Here, the unit cell from the condensed phase was chosen to be the cluster for representing the graphical partitioning. In all cases, it was also found that the basis-set extrapolation results appear to be transferable across DFT functionals with about a tenth of the computational cost and less than 0.5 kcal/mol lattice energy error. For the DFT functional extrapolations, the hybrid functionals PBE0 and B3LYP were well described at costs commensurate with PBE and revTPSS. Overall, this approach is demonstrated to perform well for both homogeneous and heterogeneous systems.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.9b01089.

The supplementary information document provides complementary figures and analyses to those presented in the main paper. Specifically, the accuracy and scaling of periodic systems is inspected in greater detail and presents a backdrop for the current publication. A detailed study on the structural library used in the main paper is presented and complements the discussion in the main paper. A detailed discussion of the dependence of error on choice of rank, $\mathcal{R}$, is carried out and complements Sections IV and V in the main paper. Lastly, we also present a selection of calculations from the paper where we consider the use of conformational energy errors in addition to the error in lattice energy. We see these two error analysis techniques to be complementary and support all outcomes presented in the paper (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Srinivasan S. Iyengar − *Department of Chemistry and Department of Physics, Indiana University, Bloomington, Indiana 47405, United States;* orcid.org/0000-0001-6526-2907; Email: iyengar@indiana.edu

Author

Timothy C. Ricard − *Department of Chemistry and Department of Physics, Indiana University, Bloomington, Indiana 47405, United States*

Complete contact information is available at:

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

https://pubs.acs.org/10.1021/acs.jctc.9b01089

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by the National Science Foundation grant NSF CHE-1665336 to S.S.I.

■ REFERENCES

(1) Hammer, B.; Nørskov, J. Theoretical surface science and catalysis - calculations and concepts. Impact of Surface Science on Catalysis. *Advances in Catalysis*; Academic Press, 2000; Vol. *45*; p 71.
(2) Wang, Z.-T.; Wang, Y.-G.; Mu, R.; Yoon, Y.; Dahal, A.; Schenter, G. K.; Glezakou, V.-A.; Rousseau, R.; Lyubinetsky, I.; Dohnálek, Z. Probing equilibrium of molecular and deprotonated water on TiO_2 (110). *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 1801.
(3) Ashwell, A. P.; Lin, W.; Hofman, M. S.; Yang, Y.; Ratner, M. A.; Koel, B. E.; Schatz, G. C. Hydrogenation of CO to Methanol on Ni (110) through Subsurface Hydrogen. *J. Am. Chem. Soc.* **2017**, *139*, 17582.
(4) Lin, W.; Schatz, G. C. Mechanisms of Formaldehyde and C_2 Formation from Methylene Reacting with CO_2 Adsorbed on Ni (110). *J. Phys. Chem. C* **2018**, *122*, 13827.
(5) Breslow, R. Hydrophobic Effects On Simple Organic-Reactions In Water. *Acc. Chem. Res.* **1991**, *24*, 159.
(6) Narayan, S.; Muldoon, J.; Finn, M. G.; Fokin, V. V.; Kolb, H. C.; Sharpless, K. B. "On Water": Unique reactivity of organic compounds in aqueous suspension. *Angew. Chem., Int. Ed.* **2005**, *44*, 3275.
(7) Davis, J. G.; Rankin, B. M.; Gierszal, K. P.; Ben-Amotz, D. On The Cooperative Formation Of Non-Hydrogen-Bonded Water At Molecular Hydrophobic Interfaces. *Nat. Chem.* **2013**, *5*, 796.
(8) Jung, Y.; Marcus, R. A. On The Theory Of Organic Catalysis On Water. *J. Am. Chem. Soc.* **2007**, *129*, 5492.
(9) Karhan, K.; Khaliullin, R. Z.; Kühne, T. D. On The Role Of Interfacial Hydrogen Bonds In "On-Water" Catalysis. *J. Chem. Phys.* **2014**, *141*, 22D528.
(10) Yan, X.; Cheng, H.; Zare, R. N. Two-phase reactions in microdroplets without the use of phase-transfer catalysts. *Angew. Chem., Int. Ed.* **2017**, *56*, 3562.
(11) Kudin, K. N.; Scuseria, G. E. Linear-scaling density-functional theory with gaussian orbitals and periodic boundary conditions: Efficient evaluation of energy and forces via the fast multipole method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2000**, *61*, 16440.
(12) Kudin, K. N.; Scuseria, G. E. Range definitions for Gaussian-type charge distributions in fast multipole methods. *J. Chem. Phys.* **1999**, *111*, 2351.
(13) Kudin, K. N.; Scuseria, G. E.; Martin, R. L. Hybrid Density-Functional Theory and the Insulating Gap of UO_2. *Phys. Rev. Lett.* **2002**, *89*, 266402.
(14) Evarestov, R. A. *Quantum Chemistry of Solids: The LCAO First Principles Treatment of Crystals*; Springer Science & Business Media, 2007; Vol. *153*.
(15) Hasnip, P. J.; Refson, K.; Probert, M. I. J.; Yates, J. R.; Clark, S. J.; Pickard, C. J. Density functional theory in the solid state. *Philos. Trans. R. Soc., A* **2014**, *372*, 20130270.
(16) Li, J.; Sode, O.; Voth, G. A.; Hirata, S. A solid−solid phase transition in carbon dioxide at high pressures and intermediate temperatures. *Nat. Commun.* **2013**, *4*, 2647.
(17) Towler, M. D.; Zupan, A.; Causà, M. Density functional theory in periodic systems using local Gaussian basis sets. *Comput. Phys. Commun.* **1996**, *98*, 181.
(18) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S.; Sugiki, S.; Sekino, H. Fast electron correlation methods for molecular clusters in the ground and excited states. *Mol. Phys.* **2005**, *103*, 2255.
(19) Bylaska, E. J.; Tsemekhman, K.; Baden, S. B.; Weare, J. H.; Jonsson, H. Parallel Implementation of Γ-Point Pseudopotential Plane-Wave DFT with Exact Exchange. *J. Comput. Chem.* **2011**, *32*, 54.
(20) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. The Nature of the Hydrated Excess Proton in Water. *Nature* **1999**, *397*, 601.
(21) Togo, A.; Oba, F.; Tanaka, I. First-principles calculations of the ferroelastic transition between rutile-type and $CaCl_2$-type SiO_2 at high pressures. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *78*, 134106.
(22) Radin, M. D.; Siegel, D. J. Charge transport in lithium peroxide: relevance for rechargeable metal−air batteries. *Energy Environ. Sci.* **2013**, *6*, 2370.
(23) Lyons, J. L.; Janotti, A.; Van de Walle, C. Effects of carbon on the electrical and optical properties of InN, GaN, and AlN. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 035204.
(24) Motta, C.; El-Mellouhi, F.; Kais, S.; Tabet, N.; Alharbi, F.; Sanvito, S. Revealing the role of organic cations in hybrid halide perovskite $CH_3NH_3PbI_3$. *Nat. Commun.* **2015**, *6*, 7026.
(25) Vidossich, P.; Lledós, A.; Ujaque, G. First-Principles Molecular Dynamics Studies of Organometallic Complexes and Homogeneous Catalytic Processes. *Acc. Chem. Res.* **2016**, *49*, 1271.
(26) Tran, V.; Soklaski, R.; Liang, Y.; Yang, L. Layer-controlled band gap and anisotropic excitons in few-layer black phosphorus. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 235319.
(27) Gerber, R. B.; Varner, M. E.; Hammerich, A. D.; Riikonen, S.; Murdachaew, G.; Shemesh, D.; Finlayson-Pitts, B. J. Computational Studies Of Atmospherically-Relevant Chemical Reactions In Water Clusters And On Liquid Water And Ice Surfaces. *Acc. Chem. Res.* **2015**, *48*, 399.
(28) Lan, J.; Hutter, J.; Iannuzzi, M. First-Principles Simulations of an Aqueous CO/Pt(111) Interface. *J. Phys. Chem. C* **2018**, *122*, 24068.
(29) Zade, S. S.; Zamoshchik, N.; Bendikov, M. From short conjugated oligomers to conjugated polymers. Lessons from studies on long conjugated oligomers. *Acc. Chem. Res.* **2010**, *44*, 14.
(30) Huzak, M.; Deleuze, M. S.; Hajgató, B. Half-metallicity and spin-contamination of the electronic ground state of graphene nanoribbons and related systems: An impossible compromise? *J. Chem. Phys.* **2011**, *135*, 104704.
(31) André, J.-M. *Electronic Structure of Polymers and Molecular Crystals*; Springer Science & Business Media, 2013; Vol. *9*.
(32) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
(33) Peverati, R.; Truhlar, D. The Quest for a Universal Density Functional: The Accuracy of Density Functionals Across a Broad Spectrum of Databases in Chemistry and Physics. *Philos. Trans. R. Soc., A* **2014**, *372*, 10120476.
(34) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315.
(35) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006**, *124*, 034108.
(36) Goerigk, L.; Grimme, S. Double-hybrid density functionals. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 576.
(37) Najibi, A.; Goerigk, L. A comprehensive assessment of the effectiveness of orbital optimization in double-hybrid density functionals in the treatment of thermochemistry, kinetics, and noncovalent interactions. *J. Phys. Chem. A* **2018**, *122*, 5610.
(38) DiStasio, R. A., Jr.; Santra, B.; Li, Z.; Wu, X.; Car, R. The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water. *J. Chem. Phys.* **2014**, *141*, 084502.
(39) Mandal, S.; Debnath, J.; Meyer, B.; Nair, N. N. Enhanced sampling and free energy calculations with hybrid functionals and plane waves for chemical reactions. *J. Chem. Phys.* **2018**, *149*, 144113.
(40) Li, X.; Iyengar, S. S. Quantum Wavepacket Ab Initio Molecular Dynamics for Extended Systems. *J. Phys. Chem. A* **2011**, *115*, 6269.
(41) Janesko, B. G.; Henderson, T. M.; Scuseria, G. E. Screened hybrid density functionals for solid-state chemistry and physics. *Phys. Chem. Chem. Phys.* **2009**, *11*, 443.

(42) Lucero, M. J.; Henderson, T. M.; Scuseria, G. E. Improved semiconductor lattice parameters and band gaps from a middle-range screened hybrid exchange functional. *J. Phys.: Condens. Matter* **2012**, *24*, 145504.
(43) Peverati, R.; Truhlar, D. G. Screened-exchange density functionals with broad accuracy for chemistry and solid-state physics. *Phys. Chem. Chem. Phys.* **2012**, *14*, 16187.
(44) Bylaska, E. J. Plane-Wave DFT Methods for Chemistry. *Annul Reports in Computational Chemistry*; Elsevier, 2017; Vol. *13*, p 185.
(45) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter* **2009**, *21*, 395502.
(46) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953.
(47) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758.
(48) Kleinman, L.; Bylander, D. M. Efficacious Form for Model Pseudopotentials. *Phys. Rev. Lett.* **1982**, *48*, 1425.
(49) Hamann, D. R.; Schlüter, M.; Chiang, C. Norm-Conserving Pseudopotentials. *Phys. Rev. Lett.* **1979**, *43*, 1494.
(50) Carrier, P.; Rohra, S.; Görling, A. General treatment of the singularities in Hartree-Fock and exact-exchange Kohn-Sham methods for solids. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2007**, *75*, 205126.
(51) Broqvist, P.; Alkauskas, A.; Pasquarello, A. Hybrid-functional calculations with plane-wave basis sets: Effect of singularity correction on total energies, energy eigenvalues, and defect energy levels. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2009**, *80*, 085114.
(52) Nguyen, H.-V.; de Gironcoli, S. Efficient calculation of exact exchange and RPA correlation energies in the adiabatic-connection fluctuation-dissipation theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2009**, *79*, 205114.
(53) Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71*, 1085.
(54) White, C. A.; Head-Gordon, M. Derivation and efficient implementation of the fast multipole method. *J. Chem. Phys.* **1994**, *101*, 6593.
(55) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. Achieving Linear Scaling for the Electronic Quantum Coulomb Problem. *Science* **1996**, *271*, 51.
(56) Li, X.-P.; Nunes, R. W.; Vanderbilt, D. Density-matrix electronic-structure method with linear system-size scaling. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 10891.
(57) Allen, M. P.; Tildesley, D. *Computer Simulation of Liquids*; Oxford Science Publications, New York, 1987.
(58) Becke, A. D. Density functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648.
(59) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange−correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51.
(60) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119*, 12129.
(61) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158.
(62) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098.
(63) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(64) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Constantin, L. A.; Sun, J. Workhorse semilocal density functional for condensed matter physics and quantum chemistry. *Phys. Rev. Lett.* **2009**, *103*, 026403.
(65) 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. A test for Diels- Alder reactions and $Pt(P(t\text{-}Bu)_3)_2$+ H_2 oxidative addition. *J. Phys. Chem.* **1996**, *100*, 19357.
(66) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. The ONIOM Method: Its Foundation and Applications to Metalloenzymes and Photobiology. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 327.
(67) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678.
(68) Varandas, A. n. J. C.; Murrell, J. N. A many-body expansion of polyatomic potential energy surfaces: application to H_n systems. *Faraday Discuss.* **1977**, *62*, 92.
(69) Murrell, J.; Carter, S.; Farantos, S.; Huxley, P.; Varandas, A. *Molecular Potential Energy Functions*; Wiley, New York, 1984.
(70) Varandas, A. J. C.; Pais, A. A. C. C. A realistic double many-body expansion (DMBE) potential energy surface for ground-state O_3 from a multiproperty fit to ab initio calculations, and to experimental spectroscopic, inelastic scattering, and kinetic isotope thermal rate data. *Mol. Phys.* **1988**, *65*, 843.
(71) Xantheas, S. S. Ab Initio Studies of Cyclic Water Clusters $(H_2O)_n$, n = 16. II. Analysis of Manybody Interactions. *J. Chem. Phys.* **1994**, *100*, 7523.
(72) Xantheas, S. S. Ab Initio Studies of Cyclic Water Clusters $(H_2O)_n$, n = 16. III. Comparison of Density Functional with MP2 Results. *J. Chem. Phys.* **1995**, *102*, 4505.
(73) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, Revision B.01; Gaussian Inc: Wallingford CT 2016.
(74) 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.
(75) 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.
(76) Haycraft, C.; Li, J.; Iyengar, S. S. "On-the-fly" Ab initio molecular dynamics with coupled cluster accuracy. *J. Chem. Theory Comput.* **2017**, *13*, 1887.
(77) Ricard, T. C.; Haycraft, C.; Iyengar, S. S. Adaptive, geometric networks for efficient coarse-grained ab initio molecular dynamics with post-Hartree-Fock accuracy. *J. Chem. Theory Comput.* **2018**, *14*, 2852.
(78) Ricard, T. C.; Iyengar, S. S. Efficiently capturing weak interactions in ab initio molecular dynamics through "on-the-fly" basis set extrapolation. *J. Chem. Theory Comput.* **2018**, *14*, 5535.
(79) Kumar, A.; Iyengar, S. S. Fragment-based electronic structure for potential energy surfaces using a superposition of fragmentation topologies. *J. Chem. Theory Comput.* **2019**, *15*, 5769.
(80) Ricard, T. C.; Kumar, A.; Iyengar, S. S. Embedded, graph-theoretically defined many-body approximations for wavefunction-in-DFT and DFT-in-DFT: applications to gas- and condensed-phase AIMD, and potential surfaces for quantum nuclear effects. *Int. J. Quantum Chem.* **2020**, DOI: 10.1002/qua.26244.
(81) Hopkins, B. W.; Tschumper, G. S. A multicentered approach to integrated QM/QM calculations. Applications to multiply hydrogen bonded systems. *J. Comput. Chem.* **2003**, *24*, 1563.

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

(82) Guo, W.; Wu, A.; Xu, X. An Extended ONIOM Method for Accurate and Efficient Geometry Optimization of Large Molecules. *Chem. Phys. Lett.* **2010**, *498*, 203.
(83) Sahu, N.; Yeole, S. D.; Gadre, S. R. Appraisal of molecular tailoring approach for large clusters. *J. Chem. Phys.* **2013**, *138*, 104101.
(84) 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.
(85) Chen, W.-K.; Fang, W.-H.; Cui, G. A multi-layer energy-based fragment method for excited states and nonadiabatic dynamics. *Phys. Chem. Chem. Phys.* **2019**, *21*, 22695.
(86) Greenwell, C.; McKinley, J. L.; Zhang, P.; Zeng, Q.; Sun, G.; Li, B.; Wen, S.; Beran, G. J. O. Overcoming the difficulties of predicting conformational polymorph energetics in molecular crystals via correlated wavefunction methods. *Chem. Sci.* **2020**, *11*, 2200.
(87) Liu, K.-Y.; Herbert, J. M. Energy-Screened Many-Body Expansion: A Practical Yet Accurate Fragmentation Method for Quantum Chemistry. *J. Chem. Theory Comput.* **2020**, *16*, 475.
(88) Demerdash, O.; Head-Gordon, T. Convergence of the many-body expansion for energy and forces for classical polarizable models in the condensed phase. *J. Chem. Theory Comput.* **2016**, *12*, 3884.
(89) Zhang, D. W.; Zhang, J. Z. H. Molecular Fractionation with Conjugate Caps for Full Quantum Mechanical Calculation of Proteinmolecule Interaction Energy. *J. Chem. Phys.* **2003**, *119*, 3599.
(90) 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.
(91) Li, S.; Li, W.; Ma, J. Generalized Energy-Based Fragmentation Approach and Its Applications to Macromolecules and Molecular Aggregates. *Accounts Chem. Res.* **2014**, *47*, 2712.
(92) Gordon, M. S.; Mullin, J. M.; Pruitt, S. R.; Roskop, L. B.; Slipchenko, L. V.; Boatz, J. A. Accurate Methods for Large Molecular Systems. *J. Phys. Chem. B* **2009**, *113*, 9646.
(93) 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.
(94) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643.
(95) Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.* **2015**, *115*, 5607.
(96) Willow, S. Y.; Salim, M. A.; Kim, K. S.; Hirata, S. Ab initio molecular dynamics of liquid water using embedded-fragment second-order many-body perturbation theory towards its accurate property prediction. *Sci. Rep.* **2015**, *5*, 14358.
(97) Liu, J.; Qi, L.-W.; Zhang, J. Z. H.; He, X. Fragment Quantum Mechanical Method for Large-Sized Ion−Water Clusters. *J. Chem. Theory Comput.* **2017**, *13*, 2021.
(98) Herbert, J. M. Fantasy versus reality in fragment-based quantum chemistry. *J. Chem. Phys.* **2019**, *151*, 170901.
(99) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many Body Expansion for Simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1.
(100) Jacobson, L. D.; Herbert, J. M. An Efficient, Fragment-Based Electronic Structure Method for Molecular Systems: Self-Consistent Polarization with Perturbative Two-Body Exchange and Dispersion. *J. Chem. Phys.* **2011**, *134*, 094118.
(101) 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.
(102) Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schutz, M.; Chan, G. K.-L. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Sci* **2014**, *345*, 640.
(103) Kerdcharoen, T.; Morokuma, K. ONIOM-XS: an extension of the ONIOM method for molecular simulation in condensed phase. *Chem. Phys. Lett.* **2002**, *355*, 257.
(104) 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.
(105) Fedorov, D. G.; Kitaura, K. Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2007**, *111*, 6904.
(106) Hirata, S. Fast Electron-Correlation Methods for Molecular Crystals: an Application to the α, $\beta(1)$, and $\beta(2)$ Modifications of Solid Formic Acid. *J. Chem. Phys.* **2008**, *129*, 204104.
(107) Kamiya, M.; Hirata, S.; Valiev, M. Fast Electron-Correlation Methods for Molecular Crystals Without Basis Set Superposition Errors. *J. Chem. Phys.* **2008**, *128*, 074103.
(108) Brorsen, K. R.; Minezawa, N.; Xu, F.; Windus, T. L.; Gordon, M. S. Fragment Molecular Orbital Molecular Dynamics with the Fully Analytic Energy Gradient. *J. Chem. Theory Comput.* **2012**, *8*, 5008.
(109) Wang, L.-W.; Zhao, Z.; Meza, J. Linear-Scaling Three-Dimensional Fragment method for Large-scale Electronic Structure Calculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 165113.
(110) 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.
(111) Han, J.; Mazack, M. J. M.; Zhang, P.; Truhlar, D. G.; Gao, J. Quantum mechanical force field for water with explicit electronic polarization. *J. Chem. Phys.* **2013**, *139*, 054503.
(112) Brorsen, K. R.; Zahariev, F.; Nakata, H.; Fedorov, D. G.; Gordon, M. S. Analytic Gradient for Density Functional Theory Based on the Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2014**, *10*, 5297.
(113) 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.
(114) Komeiji, Y.; Mochizuki, Y.; Nakano, T. Three-Body Expansion and Generalized Dynamic Fragmentation Improve the Fragment Molecular Orbital-Based Molecular Dynamics (FMO-MD). *Chem. Phys. Lett.* **2010**, *484*, 380.
(115) Liu, J.; Zhu, T.; Wang, X.; He, X.; Zhang, J. Z. H. Quantum Fragment Based ab Initio Molecular Dynamics for Proteins. *J. Chem. Theory Comput.* **2015**, *11*, 5897.
(116) Collins, M. A. Can Systematic Molecular Fragmentation Be Applied to Direct Ab Initio Molecular Dynamics? *J. Phys. Chem. A* **2016**, *120*, 9281.
(117) Liu, P.; Li, W.; Kan, Z.; Sun, H.; Ma, J. Factor Analysis of Conformations and NMR Signals of Rotaxanes: AIMD and Polarizable MD Simulations. *J. Phys. Chem. A* **2016**, *120*, 490.
(118) Liu, J.; Rana, B.; Liu, K.-Y.; Herbert, J. M. Variational Formulation of the Generalized Many-Body Expansion with Self-Consistent Charge Embedding: Simple and Correct Analytic Energy Gradient for Fragment-Based ab Initio Molecular Dynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 3877.
(119) König, C.; Christiansen, O. Linear-scaling generation of potential energy surfaces using a double incremental expansion. *J. Chem. Phys.* **2016**, *145*, 064105.
(120) Sode, O.; Keçeli, M.; Hirata, S.; Yagi, K. Coupled-cluster and many-body perturbation study of energies, structures, and phonon dispersions of solid hydrogen fluoride. *Int. J. Quantum Chem.* **2009**, *109*, 1928.
(121) Bygrave, P. J.; Allan, N. L.; Manby, F. R. The embedded many-body expansion for energetics of molecular crystals. *J. Chem. Phys.* **2012**, *137*, 164102.
(122) Giese, T. J.; York, D. M. Ambient-potential composite Ewald method for ab initio quantum mechanical/molecular mechanical molecular dynamics simulation. *J. Chem. Theory Comput.* **2016**, *12*, 2611.
(123) Červinka, C.; Fulem, M.; Růžička, K. CCSD (T)/CBS fragment-based calculations of lattice energy of molecular crystals. *J. Chem. Phys.* **2016**, *144*, 064505.
(124) Liu, J.; He, X.; Zhang, J. Z. H. Structure of liquid water a dynamical mixture of tetrahedral and ring-and-chain like structures. *Phys. Chem. Chem. Phys.* **2017**, *19*, 11931.

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

(125) Liu, J.; Sun, H.; Glover, W. J.; He, X. Prediction of excited-state properties of oligoacene crystals using fragment-based quantum mechanical method. *J. Phys. Chem. A* **2019**, *123*, 5407.
(126) Hartman, J. D.; Balaji, A.; Beran, G. J. O. Improved Electrostatic Embedding for Fragment-Based Chemical Shift Calculations in Molecular Crystals. *J. Chem. Theory Comput.* **2017**, *13*, 6043.
(127) Borca, C. H.; Bakr, B. W.; Burns, L. A.; Sherrill, C. D. CrystaLattE: Automated computation of lattice energies of organic crystals exploiting the many-body expansion to achieve dual-level parallelism. *J. Chem. Phys.* **2019**, *151*, 144103.
(128) Fang, T.; Li, Y.; Li, S. Generalized energy-based fragmentation approach for modeling condensed phase systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *7*, No. e1297.
(129) Sontising, W.; Beran, G. J. O. Theoretical assessment of the structure and stability of the λ phase of nitrogen. *Phys. Rev. Mater.* **2019**, *3*, 095002.
(130) 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.
(131) Beran, G. J. O. A new era for ab initio molecular crystal lattice energy prediction. *Angew. Chem., Int. Ed.* **2015**, *54*, 396.
(132) Demerdash, O.; Mao, Y.; Liu, T.; Head-Gordon, M.; Head-Gordon, T. Assessing many-body contributions to intermolecular interactions of the AMOEBA force field using energy decomposition analysis of electronic structure calculations. *J. Chem. Phys.* **2017**, *147*, 161721.
(133) Ghosh, D. Hybrid Equation-of-Motion Coupled-Cluster/Effective Fragment Potential Method: A Route toward Understanding Photoprocesses in the Condensed Phase. *J. Phys. Chem. A* **2017**, *121*, 741.
(134) Hermann, A.; Schwerdtfeger, P. Ground-state properties of crystalline ice from periodic Hartree-Fock calculations and a coupled-cluster-based many-body decomposition of the correlation energy. *Phys. Rev. Lett.* **2008**, *101*, 183005.
(135) Bludský, O.; Rubeš, M.; Soldán, P. Ab initio investigation of intermolecular interactions in solid benzene. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 092103.
(136) Tsuzuki, S.; Orita, H.; Honda, K.; Mikami, M. First-principles lattice energy calculation of urea and hexamine crystals by a combination of periodic DFT and MP2 two-body interaction energy calculations. *J. Phys. Chem. B* **2010**, *114*, 6799.
(137) Meitei, O. R.; Hesselmann, A. On the Stability of Cyclophane Derivates Using a Molecular Fragmentation Method. *ChemPhysChem* **2016**, *17*, 3863.
(138) Dolgonos, G. A.; Loboda, O. A.; Boese, A. D. Development of Embedded and Performance of Density Functional Methods for Molecular Crystals. *J. Phys. Chem. A* **2018**, *122*, 708.
(139) De Jong, B. P. H. K.; Wilson, J. E.; Neilson, G. W.; Buckingham, A. D. Hydrophobic Hydration Of Methane. *Mol. Phys.* **1997**, *91*, 99.
(140) Gajewski, J. J. The Claisen rearrangement. Response to solvents and substituents: The case for both hydrophobic and hydrogen bond acceleration in water and for a variable transition state. *Acc. Chem. Res.* **1997**, *30*, 219.
(141) Huang, L.; Massa, L.; Karle, J. Kernel energy method: Application to DNA. *Biochem* **2005**, *44*, 16747.
(142) Stoll, H.; Paulus, B.; Fulde, P. On the accuracy of correlation-energy expansions in terms of local increments. *J. Chem. Phys.* **2005**, *123*, 144108.
(143) Paulus, B. The method of increments - a wavefunction-based ab initio correlation method for solids. *Phys. Rep.* **2006**, *428*, 1.
(144) 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.
(145) Beran, G. J. O.; Hirata, S. Fragment and Localized Orbital Methods in Electronic Structure Theory. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7559.
(146) Collins, M. A. Systematic Fragmentation of Large Molecules by Annihilation. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7744.
(147) Beran, G. J. O.; Wen, S.; Nanda, K.; Huang, Y.; Heit, Y. Accurate and Robust Molecular Crystal Modeling Using Fragment-Based Electronic Structure Methods. In *Prediction and Calculation of Crystal Structures: Methods and Applications*; Atahan-Evrenk, S., Aspuru-Guzik, A., Eds.; Topics in Current Chemistry-Series, 2014; Vol. *345*, p 59.
(148) Červinka, C.; Beran, G. J. O. Ab initio prediction of the polymorph phase diagram for crystalline methanol. *Chem. Sci.* **2018**, *9*, 4622.
(149) Sexton, T. M.; Tschumper, G. S. 2-body:Many-body QM:QM study of structures, energetics, and vibrational frequencies for microhydrated halide ions. *Mol. Phys.* **2019**, *117*, 1413.
(150) Mecke, K. R.; Wagner, H. Euler characteristic and related measures for random geometric sets. *J. Stat. Phys.* **1991**, *64*, 843.
(151) Dey, T. K.; Shah, N. R. On the number of simplicial complexes in R^d. *Comput. Geom.* **1997**, *8*, 267.
(152) Adams, C. C.; Franzosa, R. D. *Introduction to Topology: Pure and Applied*; Pearson Prentice Hall: Upper Saddle River, 2008.
(153) Berger, M.; Pansu, P.; Berry, J.-P.; Saint-Raymond, X. Affine spaces. *Problems in Geometry*; Springer, 1984; p 11.
(154) Björklund, A.; Husfeldt, T.; Koivisto, M. Set partitioning via inclusion-exclusion. *SIAM J. Comput.* **2009**, *39*, 546.
(155) Nanda, K. D.; Beran, G. J. O. Prediction of organic molecular crystal geometries from MP2-level fragment quantum mechanical/molecular mechanical calculations. *J. Chem. Phys.* **2012**, *137*, 174106.
(156) Huang, Y.; Beran, G. J. O. Reliable prediction of three-body intermolecular interactions using dispersion-corrected second-order Moller-Plesset perturbation theory. *J. Chem. Phys.* **2015**, *143*, 044113.
(157) Beran, G. J. O.; Hartman, J. D.; Heit, Y. N. Predicting Molecular Crystal Properties from First Principles: Finite Temperature Thermochemistry to NMR Crystallography. *Acc. Chem. Res.* **2016**, *49*, 2501.
(158) Beran, G. J. O. Modeling Polymorphic Molecular Crystals with Electronic Structure Theory. *Chem. Rev.* **2016**, *116*, 5567.
(159) Hartman, J. D.; Kudla, R. A.; Day, G. M.; Mueller, L. J.; Beran, G. J. O. Benchmark fragment-based ^{1}H, ^{1}3C, ^{1}5N and ^{1}7O chemical shift predictions in molecular crystals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 21686.
(160) Červinka, C.; Beran, G. J. O. Towards reliable ab initio sublimation pressures for organic molecular crystals - are we there yet? *Phys. Chem. Chem. Phys.* **2019**, *21*, 14799.
(161) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73.
(162) Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E., III; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.; Gonthier, J. F.; James, A. M.; McAlexander, H. R.; Kumar, A.; Saitow, M.; Wang, X.; Pritchard, B. P.; Verma, P.; Schaefer, H. F., III; Patkowski, K.; King, R. A.; Valeev, E. F.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. PSI4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Theor. Comput. Chem.* **2017**, *13*, 3185.
(163) Ozaki, T.; Kino, H.; Yu, J.; Han, M.; Ohfuchi, M.; Ishii, F.; Sawada, K.; Ohwaki, T.; Weng, H.; Toyoda, M.; Okuno, Y.; Perez, R.; Bell, P.; Duy, T. V. T.; Xiao, Y.; Ito, A.; Terakura, K. *Users Manual of OpenMX Ver. 3.8*, 2016.
(164) Moreno, J.; Soler, J. M. Optimal meshes for integrals in real-and reciprocal-space unit cells. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *45*, 13891.
(165) Suhai, S.; Bagus, P. S.; Ladik, J. An error analysis for Hartree-Fock crystal orbital calculations. *Chem. Phys.* **1982**, *68*, 467.
(166) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **1999**, *151*, 283.
(167) Liang, W.; Baer, R.; Saravanan, C.; Shao, Y.; Bell, A. T.; Head-Gordon, M. Fast methods for resumming matrix polynomials and Chebyshev matrix polynomials. *J. Comput. Phys.* **2004**, *194*, 575.
(168) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.;

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812

O'Neill, D. P.; DiStasio, R. A., Jr.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Yeh Lin, C.; Van Voorhis, T.; Hung Chien, S.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Min Rhee, Y.; Ritchie, J.; Rosta, E.; David Sherrill, C.; Simmonett, A. C.; Subotnik, J. E.; Lee Woodcock III, H., III; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer III, H. F., III; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172.

(169) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. An efficient self-consistent field method for large systems of weakly interacting components. *J. Chem. Phys.* **2006**, *124*, 204105.

(170) Sun, Z.; Zhu, T.; Wang, X.; Mei, Y.; Zhang, J. Z. H. Optimization of convergence criteria for fragmentation methods. *Chem. Phys. Lett.* **2017**, *687*, 163.

(171) Knuth, D. E. *The Art of Computer Programming*; Addison-Wesley: Reading, Massachusetts, 1973.

(172) Manber, U. *Introduction to Algorithms: A Creative Approach*; Pearson Education: New Jersey, 1989.

(173) Bates, D. M.; Tschumper, G. S. CCSD (T) complete basis set limit relative energies for low-lying water hexamer structures. *J. Phys. Chem. A* **2009**, *113*, 3555.

(174) Wang, Y.; Babin, V.; Bowman, J. M.; Paesani, F. The water hexamer: cage, prism, or both. Full dimensional quantum simulations say both. *J. Am. Chem. Soc.* **2012**, *134*, 11116.

(175) Pruitt, S. R.; Steinmann, C.; Jensen, J. H.; Gordon, M. S. Fully Integrated Effective Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2013**, *9*, 2235.

(176) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Effective Fragment Molecular Orbital Method: A Merger of the Effective Fragment Potential and Fragment Molecular Orbital Methods. *J. Phys. Chem. A* **2010**, *114*, 8705.

(177) Richard, R. M.; Herbert, J. M. Many-Body Expansion with Overlapping Fragments: Analysis of Two Approaches. *J. Chem. Theory Comput.* **2013**, *9*, 1408.

(178) Hua, S.; Hua, W.; Li, S. An Efficient Implementation of the Generalized Energy-Based Fragmentation Approach for General Large Molecules. *J. Phys. Chem. A* **2010**, *114*, 8126.

(179) Hopkins, B. W.; Tschumper, G. S. Integrated quantum mechanical approaches for extended π systems: Multicentered QM/QM studies of the cyanogen and diacetylene trimers. *Chem. Phys. Lett.* **2005**, *407*, 362.

(180) Richard, R. M.; Lao, K. U.; Herbert, J. M. Aiming for Benchmark Accuracy with the Many-Body Expansion. *Acc. Chem. Res.* **2014**, *47*, 2828.

(181) Ouyang, J. F.; Bettens, R. P. A. When are Many-Body Effects Significant? *J. Chem. Theory Comput.* **2016**, *12*, 5860.

(182) Lao, K. U.; Liu, K.-Y.; Richard, R. M.; Herbert, J. M. Understanding the many-body expansion for large systems. II. Accuracy considerations. *J. Chem. Phys.* **2016**, *144*, 164105.

(183) Ouyang, J. F.; Bettens, R. P. A. Many-Body Basis Set Superposition Effect. *J. Chem. Theory Comput.* **2015**, *11*, 5132.

(184) Paier, J.; Marsman, M.; Hummer, K.; Kresse, G.; Gerber, I. C.; Ángyán, J. G. Screened hybrid density functionals applied to solids. *J. Chem. Phys.* **2006**, *124*, 154709.

(185) Paier, J.; Marsman, M.; Kresse, G. Why does the B3LYP hybrid functional fail for metals? *J. Chem. Phys.* **2007**, *127*, 024103.

(186) Guidon, M.; Hutter, J.; VandeVondele, J. Robust Periodic Hartree- Fock Exchange for Large-Scale Simulations Using Gaussian Basis Sets. *J. Chem. Theory Comput.* **2009**, *5*, 3010.

(187) Holzwarth, N. A. W.; Xu, X. Analysis of numerical methods for evaluating the Fock exchange integral in a plane-wave basis. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2011**, *84*, 113102.

(188) Hu, W.; Lin, L.; Banerjee, A. S.; Vecharynski, E.; Yang, C. Adaptively Compressed Exchange Operator for Large-Scale Hybrid Density Functional Calculations with Applications to the Adsorption of Water on Silicene. *J. Chem. Theory Comput.* **2017**, *13*, 1188.

(189) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289.

(190) Gillan, M. J.; Alfè, D.; Michaelides, A. Perspective: How good is DFT for water? *J. Chem. Phys.* **2016**, *144*, 130901.

(191) Perdew, J. P.; Zunger, A. Self-Interaction Correction to Density-Functional Approximations for Many-Election Systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1981**, *23*, 5048.

(192) Mori-Sanchez, P.; Cohen, A. J.; Yang, W. Localization and delocalization errors in density functional theory and implications for band-gap prediction. *Phys. Rev. Lett.* **2008**, *100*, 146401.

(193) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. Insights into current limitations of density functional theory. *Science* **2008**, *321*, 792.

(194) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Buongiorno Nardelli, M.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M.; Colonna, N.; Carnimeo, I.; Dal Corso, A.; de Gironcoli, S.; Delugas, P.; DiStasio, R. A., Jr.; Ferretti, A.; Floris, A.; Fratesi, G.; Fugallo, G.; Gebauer, R.; Gerstmann, U.; Giustino, F.; Gorni, T.; Jia, J.; Kawamura, M.; Ko, H.-Y.; Kokalj, A.; Küçükbenli, E.; Lazzeri, M.; Marsili, M.; Marzari, N.; Mauri, F.; Nguyen, N. L.; Nguyen, H.-V.; Otero-de-la-Roza, A.; Paulatto, L.; Poncé, S.; Rocca, D.; Sabatini, R.; Santra, B.; Schlipf, M.; Seitsonen, A. P.; Smogunov, A.; Timrov, I.; Thonhauser, T.; Umari, P.; Vast, N.; Wu, X.; Baroni, S. Advanced capabilities for materials modelling with QUANTUM ESPRESSO. *J. Phys.: Condens. Matter* **2017**, *29*, 465901.

(195) Tinkham, M. *Group Theory and Quantum Mechanics*; McGraw-Hill Book Company, 1964.

(196) Van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. State of the art in counterpoise theory. *Chem. Rev.* **1994**, *94*, 1873.

(197) Breslow, R.; Maitra, U.; Rideout, D. Selective Diels-Alder Reactions In Aqueous-Solutions And Suspensions. *Tetrahedron Lett.* **1983**, *24*, 1901.

(198) Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.* **1980**, *102*, 939.

(199) Jung, Y.; Marcus, R. A. Protruding Interfacial OH Groups And "On-Water" Heterogeneous Catalysis. *J. Phys.: Condens. Matter* **2010**, *22*, 284117.

(200) Sahota, N.; AbuSalim, D. I.; Wang, M. L.; Brown, C. J.; Zhang, Z.; El-Baba, T. J.; Cook, S. P.; Clemmer, D. E. A microdroplet-accelerated Biginelli reaction: mechanisms and separation of isomers using IMS-MS. *Chem. Sci.* **2019**, *10*, 4822.

https://dx.doi.org/10.1021/acs.jctc.9b01089
J. Chem. Theory Comput. 2020, 16, 4790−4812