

Capturing Weak Interactions in Surface Adsorbate Systems at Coupled Cluster Accuracy: A Graph-Theoretic Molecular Fragmentation Approach Improved through Machine Learning

Timothy C. Ricard, Xiao Zhu, and Srinivasan S. Iyengar*

Cite This: *J. Chem. Theory Comput.* 2023, 19, 8541–8556

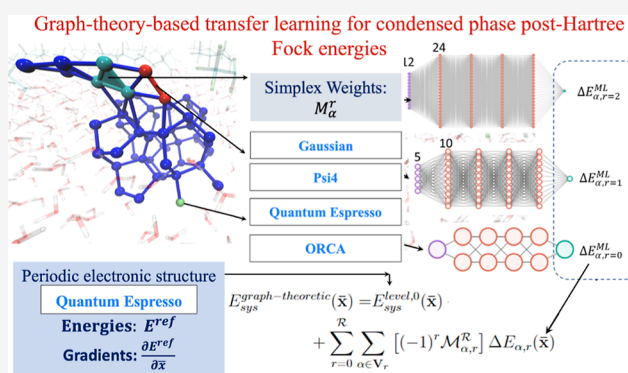
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: The accurate and efficient study of the interactions of organic matter with the surface of water is critical to a wide range of applications. For example, environmental studies have found that acidic polyfluorinated alkyl substances, especially perfluorooctanoic acid (PFOA), have spread throughout the environment and bioaccumulate into human populations residing near contaminated watersheds, leading to many systemic maladies. Thus, the study of the interactions of PFOA with water surfaces became important for the mitigation of their activity as pollutants and threats to public health. However, theoretical study of the interactions of such organic adsorbates on the surface of water, and their bulk concerted properties, often necessitates the use of ab initio methods to properly incorporate the long-range electronic properties that govern these extended systems. Notable theoretical treatments of “on-water” reactions thus far have employed hybrid DFT and semilocal DFT, but the interactions involved are weak interactions that may be best described using post-Hartree–Fock theory. Here, we aim to demonstrate the utility of a graph-theoretic approach to molecular fragmentation that accurately captures the critical “weak” interactions while maintaining an efficient ab initio treatment of the long-range periodic interactions that underpin the physics of extended systems. We apply this graph-theoretical treatment to study PFOA on the surface of water as a model system for the study of weak interactions seen in the wide range of surface interactions and reactions. The approach divides a system into a set of vertices, that are then connected through edges, faces, and higher order graph theoretic objects known as simplexes, to represent a collection of locally interacting subsystems. These subsystems are then used to construct ab initio molecular dynamics simulations and for computing multidimensional potential energy surfaces. To further improve the computational efficiency of our graph theoretic fragmentation method, we use a recently developed transfer learning protocol to construct the full system potential energy from a family of neural networks each designed to accurately model the behavior of individual simplexes. We use a unique multidimensional clustering algorithm, based on the *k*-means clustering methodology, to define our training space for each separate simplex. These models are used to extrapolate the energies for molecular dynamics trajectories at PFOA water interfaces, at less than one-tenth the cost as compared to a regular molecular fragmentation-based dynamics calculation with excellent agreement with couple cluster level of full system potential energies.



1. INTRODUCTION

Theoretical study of the interactions of organic adsorbates on the surface of water, including (a) local electronic effects arising from weak molecular interactions and (b) associated long-range electrostatic effects, is a critical challenge for ab initio theory and materials modeling. Notable theoretical treatments of “on-water” electronic interactions that are thought to accelerate organic reactions^{1–13} thus far have primarily employed hybrid density functional theory (DFT)^{13,14} and semilocal DFT.^{15–17} However, the interactions involved here are weak interactions that are best described using post-Hartree–Fock methods. Despite substantial recent progress in DFT functional development,^{18–20} multiple

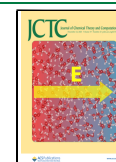
challenges remain^{18,21–29} in the treatment of weak interactions. Critically most periodic calculations are restricted to pure (semilocal) DFT due to the severe cost of including Hartree–Fock exchange^{29–35} and post-Hartree–Fock methods.^{36,37} These restrictions become especially critical when considering the calculation of ab initio molecular dynamics trajec-

Received: August 30, 2023

Revised: November 6, 2023

Accepted: November 10, 2023

Published: November 29, 2023



ries^{22,35,38} and quantum nuclear studies³⁹ where gradients are also needed for a large number of individual structures. Pure (semilocal) density functionals also suffer from the so-called self-interaction of electrons^{23–26,40} which is reduced in nonlocal functionals and is not a concern in post-Hartree–Fock methods. The accurate study of weak interactions by DFT methods^{41,42} is partially achieved by the use of empirical dispersion corrections that depend on the nuclear–nuclear interaction terms^{43,44} or by using nonlocal van der Waals (vdW) functionals.^{45,46} Recent developments have allowed lower order scaling hybrid DFT calculations in the condensed phase.^{47–49} Furthermore, large basis sets (triple- ζ with polarization and diffused basis functions) are additionally necessary to treat such interactions, and these are generally difficult to use within an AIMD and condensed phase formalism.

In this publication, we present an approach⁵⁰ that will accurately capture both the critical “weak” interactions while maintaining an efficient *ab initio* treatment of the long-range, periodic interactions which underpin the physics of extended systems. This approach treats the full target periodic system with an affordable reference level of theory, which is then perturbatively corrected to incorporate higher quality treatment of short-range interactions using a graph-theoretic description of many-body approximation (MBE). For these condensed phase systems, the long-range extended interactions are captured within a periodic calculation using a semilocal DFT treatment. Then corrections are added which capture the shorter range many-body^{51–59} expansions including post-Hartree–Fock correlation. These many-body corrections are constructed using a graphical representation^{50,60–71} of the system which allows for easy tailoring of the correction in regard to the range and order of the expansion. This approach provides both the overall energy and gradients necessary for the computation and analysis of AIMD trajectories^{60–64,68} (and potential energy surfaces^{65,67}) obtained from multiple independent electronic structure calculations. Furthermore, our graph theoretical fragmentation procedure is integrated here with an efficient neural network scheme to further enhance the efficiency for the study of condensed phase correlation problems. This machine-learning interface further accelerates the needed fragment calculations within the graph-theoretic fragmentation formalism. Machine learning methods^{72–74} have been used by many investigators to accelerate the computation of potential surfaces in small molecular systems.^{75–88} Here, we combined neural networks with a geometric scheme known as k-means clustering^{89,90} to also optimize the amount of training data needed to compute energies for the post-Hartree–Fock treatment of larger clusters.

Our Python-based driver⁶⁸ optimizes the parallel^{91,92} nature of our method and allows the use of a range of electronic structure packages in the evaluation of system energy and gradients. We apply the above graph-theory-based machine learning protocols to study a fluorocarbon surfactant, perfluorooctanoic acid (PFOA), on the surface of water as a model system for the study of weak interactions seen in a wide range of surface interactions and reactions. PFOA is a member of a large class of industrial chemicals termed perfluorinated and polyfluorinated alkyl substances (PFASs). PFASs have been utilized for water-proofing and oil-proofing of textile, leather, and paper products⁹³ and serve as nonflammable and noncorrosive mediums for sensitive materials.⁹³ These

chemicals have been shown to pose significant health and environmental threats,^{94–99} especially due to their tendency toward bioaccumulation and “biomagnification”.¹⁰⁰ PFAS molecules bioaccumulate into both wildlife and human populations^{101–104} residing near contaminated watersheds,¹⁰⁵ leading to systemic maladies^{95,106} such as thyroid¹⁰⁷ and kidney diseases,¹⁰⁸ a range of cancers,^{109–111} and infertility.^{112,113} In the early 2000s, environmental studies found that acidic PFASs, especially PFOA, had spread throughout the environment¹⁰¹ and almost all humans have trace amounts in their bodies.^{101,102,114,115} For this reason PFOA and perfluorooctanesulfonic acid,¹¹⁶ which were notably used in fire-fighting foams for fuel fires, were added to a list of regulated chemicals in drinking water in March 2021¹¹⁷ based upon recommendation of the AAAS.¹¹⁸ The cleanup of these contaminations can prove quite costly, but there has been some recent promising developments in methodologies to degrade perfluoroalkyl carboxylic acids, including PFOA.^{119,120} Thus, the study of the interactions of PFOA, and other PFASs, with water surfaces becomes important and may help develop methods for the mitigation of their activity as pollutants and threats to public health.

This article is organized as follows: in Section 2, we briefly present our graph-theoretic approach^{50,60–68} along with its machine learning enhancements toward application to condensed phase calculations.⁵⁰ The graph theoretic method is demonstrated in Section 3 by computing PFOA interactions on the surface of condensed phase water, where the molecular conformations are selected from an AIMD trajectory computed by using the graph theoretical approach. In these calculations, we extrapolate from PBE to CCSD to achieve CCSD quality structural energies. That is, all calculations computationally scale as PBE, but are tailored to provide CCSD accuracy. As higher order terms can still prove costly, we discuss the use of transfer machine learning to further accelerate these calculations. Conclusions are given in Section 4.

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

Conventional electronic structure methods incur steep algebraic scaling costs with the system size and basis set quality. These scaling costs become prohibitive when post-Hartree–Fock methods are considered, limiting the use of high quality correlation methods to small- and medium-sized clusters. In addition, their use is cost prohibitive for condensed phase systems even when considering moderate sized unit cells.^{36,37} Due to these inherent scaling costs, molecular fragmentation methods^{55–58,121–147} have been developed with the aim of limiting the computational costs by restricting the expensive calculations to smaller components of the overall systems. A number of these methods incorporate the full many-body or long-range terms through a composite^{148–151} approach where a low cost reference calculation of the full system is corrected by the many-body or fragment terms.^{139,152} In addition to their standard application to gas-phase calculations, these corrections have also been demonstrated in condensed phase systems^{143,153–159} where the full periodicity is treated with a cheaper HF or semilocal DFT and the local interactions incorporate higher quality physics.

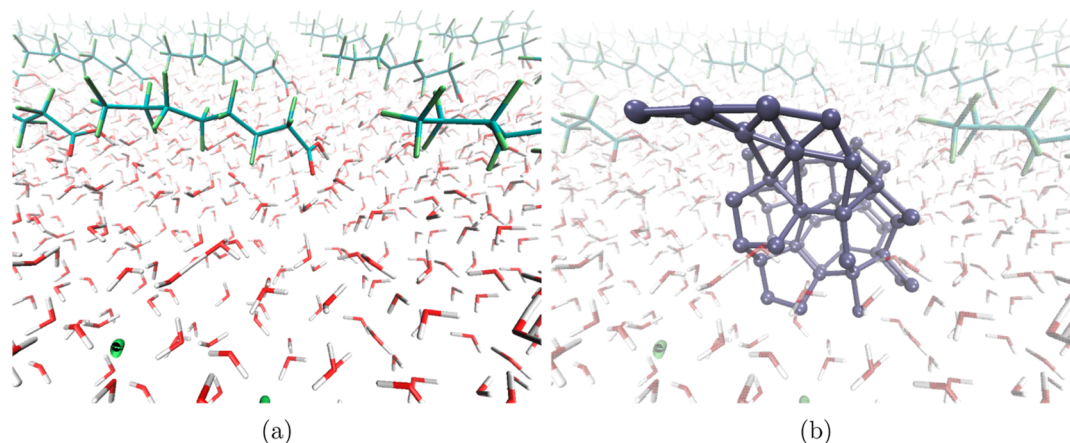


Figure 1. Graphical representation considered in eq 2 is illustrated for PFOA on a periodic slab of water. The periodicity is captured by reference energy, E^{ref} , calculated by using a plane-wave basis. Additional correlations via post-Hartree–Fock corrections are introduced by the $\Delta E_{\alpha,r}(\bar{\mathbf{x}})$ terms in eq 2 using the graphical network depiction in (b). These PFOA and water systems are partitioned into nodes of either a single water or a carbon with its associated functional groups.

Due to the advent of higher quality hardware and computational algorithms, recent applications of these methods employ Hartree–Fock or semilocal periodic DFT methods to treat long-range interactions in bulk crystals or periodic surfaces, while introducing perturbative corrections to capture shorter range correlation effects, using either hybrid DFT or post-Hartree–Fock methods, to capture short-range interactions.^{143,153–157}

Following these advances, in ref 50, we have demonstrated hybrid DFT (Rung-4 functionals) quality condensed-phase simulations for surfaces with organic adsorbate molecules and bulk systems where long-range, periodic interactions were treated with gradient-corrected DFT (Rung-2 functionals) calculations augmented by short-range perturbative MBE corrections obtained by graph theory. This was achieved by utilizing the graph-theoretic approach introduced in, refs 50,60–68,71. In this paper, we aim to expand its use to capture post-Hartree–Fock correlation effects between surfaces and adsorbates, while maintaining the scaling costs of semilocal DFT methods. This aim is supported by our method’s previous success in capturing post-Hartree–Fock correlation in both extended Lagrangian^{61–63} as well as Born–Oppenheimer^{60–63,68} (BOMD)-based ab initio molecular dynamics with DFT-computational expense. In this work, we combine these features to obtain post-Hartree–Fock quality adsorbate–surface interactions.

Our method is a composite approach that considers a reference calculation of the long-range or periodic physics of the system. This calculation is then corrected by short-range perturbative terms arising from a graphical representation. The reference calculation, E^{ref} , treats the full system using a lower scaling electronic structure method. When the full system is an isolated cluster, the reference calculation uses an atom centered basis, and when considering an extended system (as in this publication), the reference calculation is periodic DFT in a plane-wave basis. The lower level of theory and basis used on the full system is referred to as “level, 0”, which in this paper will be the semilocal DFT functional, PBE, with either a Pople style atom centered basis (for a molecular cluster) or a plane-wave basis with the Kresse–Joubert style projected augmented wave treatment^{160,161} (for a periodic system). The energy, and forces,⁶⁶ from E^{ref} is perturbatively corrected by an ONIOM-

like^{162–164} perturbative correction term generated by a graph theoretic representation of the MBE.^{51–59,152,165–167} This is done by partitioning the molecular system into a set of nodes that may consist of fragments with stable chemical properties such as a single water molecule, a methyl group, or, in the case of PFOA, fluorinated hydro-carbon fragments. Short range interactions between these nodes are constructed by including edges on the basis of a Cartesian distance criterion. The set of these nodes ($\mathbf{V}_0^{\bar{\mathbf{x}}}$) and edges ($\mathbf{V}_1^{\bar{\mathbf{x}}}$) form a graph, $\mathcal{G}_{\bar{\mathbf{x}}} \equiv \{\mathbf{V}_0^{\bar{\mathbf{x}}}, \mathbf{V}_1^{\bar{\mathbf{x}}}\}$, which represents the critical interactions for a given instantaneous structure, $\bar{\mathbf{x}}$, where two body interactions are captured. An example of how the graphical representation is defined for a single PFOA molecule inside a unit-cell on the surface of water is illustrated in Figure 1. Embedded inside of these graphical representations are triangles (rank 2 objects that capture three-body interactions), tetrahedrons (rank 3 objects that capture four-body interactions), and other higher order objects which capture higher order interaction energies. Each set of these higher order interactions forms a power set of nodes

$$\{\mathbf{V}_r^{\bar{\mathbf{x}}} | r=0 \dots \mathcal{R}\} \quad (1)$$

within the graph. Thus, this core approximation produces the energy expression

$$E_{\text{sys}}^{\text{graph-theoretic}}(\bar{\mathbf{x}}) = E^{\text{Ref}}(\bar{\mathbf{x}}) + \sum_{r=0}^{\mathcal{R}} \sum_{\alpha \in \mathbf{V}_r} [\mathcal{M}_{\alpha,r}^{\mathcal{R}}] \Delta E_{\alpha,r}(\bar{\mathbf{x}}) \quad (2)$$

where

$$\Delta E_{\alpha,r}(\bar{\mathbf{x}}) = E_{\alpha,r}^{\text{level},1}(\bar{\mathbf{x}}) - E_{\alpha,r}^{\text{level},0}(\bar{\mathbf{x}}) \quad (3)$$

Each term inside the summation in the second term on the right side in eq 2 is a correction term between level, 1 and level, 0 theory (as in ONIOM) for the α -th rank- r fragment in the system. Each node ($r=0$), edge ($r=1$), face ($r=2$), and so forth, contributes to this total correction based on their multiplicity

$$\mathcal{M}_{\alpha,r}^{\mathcal{R}} = \sum_{m=r}^{\mathcal{R}} (-1)^{m+r} p \alpha^{r,m} \quad (4)$$

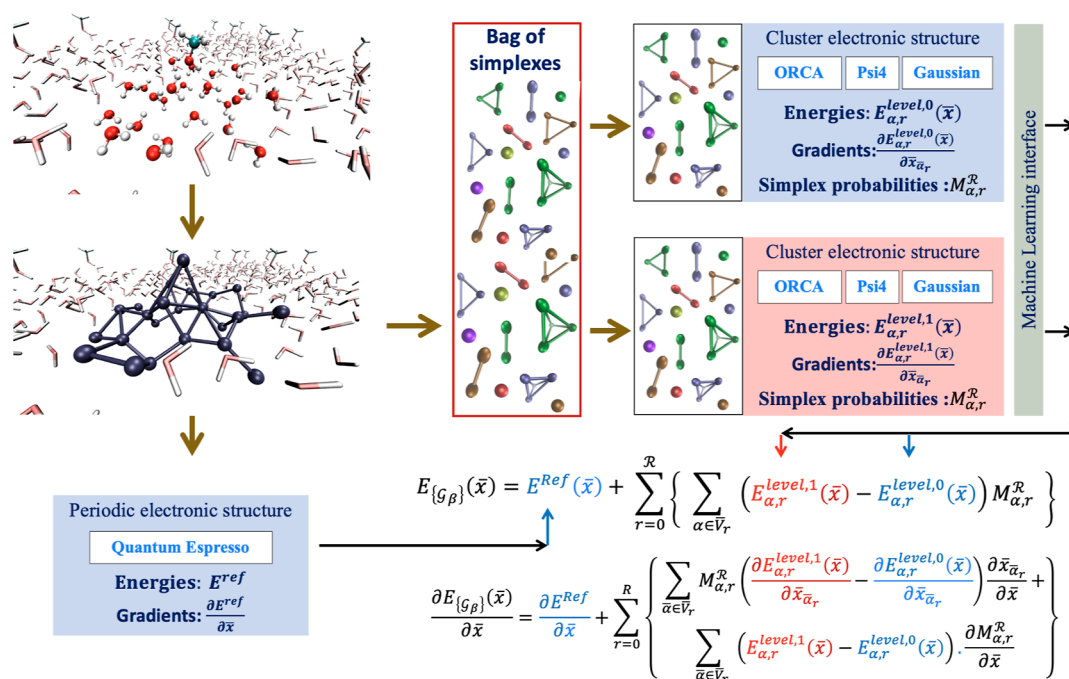


Figure 2. Summary of algorithm.

arising from their connectivity within the graphical representation. The factors, (a) $p\alpha^{r,m}$ are the number of times the α th rank- r term (in set V_r) appears in all rank- m terms (in set V_m), for $m \geq r$, and (b) consequently, $M_{\alpha,r}^R$ is the overcounting correction for the number of times the α th $(r+1)$ -body term appears in all objects of rank greater than or equal to r within a specific graphical representation. As these correction terms are inherently additive and independent, the problem of computing these energies becomes trivially parallel in practice.^{91,92} Additionally, the independent nature of these energy terms allows for the simultaneous use of multiple electronic structure packages, such as Gaussian,¹⁶⁸ ORCA,¹⁶⁹ Psi4,¹⁷⁰ and Quantum Espresso.¹⁷¹ For further information, see refs 50, 66, and 68. The entire algorithm is summarized in Figure 2.

2.1. Computing Smooth Potentials with Graph-Theory Based Molecular Fragmentation. When molecular potential surfaces are computed, as needed in AIMD, molecular fragmentation methods are known to yield discontinuities in energy and forces. For example, the fragmentation approach is dynamical and, in general, changes during AIMD. This would also be critical in constructing potential energy surfaces for quantum nuclear dynamical effects. There are two approaches that we have developed that overcome these discontinuities. In refs 65 and 67, we introduce a multitopology based fragmentation procedure, where, essentially, the overall energy (and gradients) are linear combinations of fragmentation energies. In general, one may obtain a family of graphs, numbered using the index β , such that

$$\mathcal{G}_{\bar{x},\beta} \equiv \{\mathbf{V}_0^{\bar{x};\beta}; \mathbf{V}_1^{\bar{x};\beta}\} \quad (5)$$

and

$$\{\mathbf{V}_r^{\bar{x};\beta} | r=0 \dots R\} \quad (6)$$

each providing a different representation of energy and gradients. The overall is a weighted sum of these individual graph energies, leading to

$$E_{\mathcal{R},\{\mathcal{G}_{\beta}\}}(\bar{x}) = \sum_{\beta} v_{\beta}(\bar{x}) \left[E^{\text{level},0}(\bar{x}) + \left(\sum_{r=0}^R (-1)^r \sum_{\alpha \in V_r^{\beta}} \Delta E_{\alpha,r,\beta}^{1,0}(\bar{x}) M_{\alpha,r,\beta}^R \right) \right] \\ = E^{\text{level},0}(\bar{x}) + \sum_{\beta} v_{\beta}(\bar{x}) \left(\sum_{r=0}^R (-1)^r \sum_{\alpha \in V_r^{\beta}} \Delta E_{\alpha,r,\beta}^{1,0}(\bar{x}) M_{\alpha,r,\beta}^R \right) \quad (7)$$

where the square-bracketed term $[\dots]$ in the first equation above is the same as eq 2, but now, the terms $\Delta E_{\alpha,r,\beta}^{1,0}(\bar{x})$ and $M_{\alpha,r,\beta}^R$ have $\{\mathcal{G}_{\bar{x},\beta}\}$ dependence. In refs 65 and 67, a variational algorithm is introduced to compute $\{v_{\beta}(\bar{x})\}$. Here, the energy of the system is a probabilistic sum over multiple fragmentation topologies (or graphs) since in some sense each of the graphs forces a certain locality in the electronic structure and hence may be loosely considered as “valence bond”^{172–174} constructs or “diabatic states”.^{175–183} The variational algorithm in refs 65 and 67 allows us to compute smooth potentials, but in ref 68, a numerical weighting scheme is introduced to obtain smooth energies and gradients.

2.2. Machine Learning Based Improvements to eq 2. While eq 2 has been shown to greatly reduce the computational cost in providing highly accurate post-Hartree–Fock energies and gradients, to the cost of DFT, there is a secondary, but nontrivial, computational challenge that appears as a result of using eq 2. Specifically, in ref 68, it is shown that for systems involving water, it may be necessary to include $r = 3$ (four-body) terms to accurately obtain energies and gradients. As per eq 3, energies are still necessary for these potentially larger sized clusters at the target level, level, 1, of theory. While computing the target energies for the full system

is clearly prohibitive, it may turn out that the larger fragments obtained from eq 2 are also challenging to compute.

In Figure 3, we illustrate this challenge by displaying the distribution of compute times for a range of rank-0 (node,

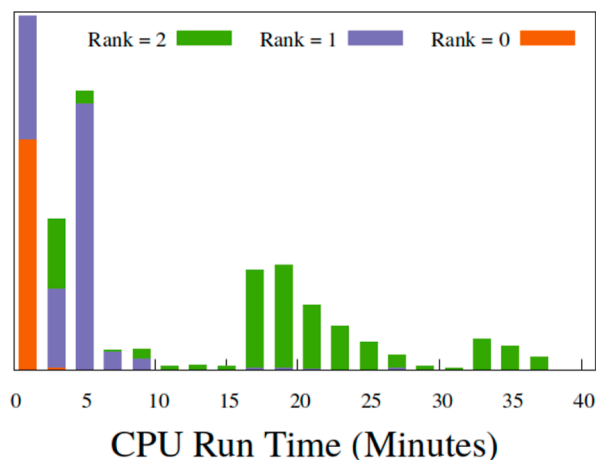


Figure 3. Distribution of run time costs to calculate $\Delta E_{\alpha,r}$ by fragment rank. Fragment computational costs grow significantly with rank; thus, as higher order corrections are added to eq 1, the efficiency of the overall approach despite parallelization may reduce drastically. It must, however, be kept in mind that full system CCSD is impossible for this system.

shown in maroon colored histograms), rank-1 (edge, displayed in blue colored histograms), and rank-2 (face, shown in green colored histograms) fragments from a single PFOA molecule on the surface of water (as shown in Figure 1). In this figure, the maroon histograms (rank-0 simplexes) are strongly peaked at the left (near zero), the blue histograms (rank-1 simplexes) are spread in regions less than 10 min, and the green histograms (rank-2 simplexes) show a range of CPU times, depending on the types of fragments and some of these could be extremely expensive to compute. Clearly, the rank-2 terms may be extremely challenging to compute and in fact, completely determine the overall cost of the calculation. Of course, it is critical to keep in mind that the full periodic system is impossible to process at the level of CCSD, which is the choice of level, 1 in Figure 3. In this section, we utilize the ideas from ref 70, to further enhance the efficiency of the

periodic post-Hartree–Fock calculations as obtained from eq 2.

Specifically, for larger clusters, we wish to replace the energies in eq 2 with machine learning estimates, that is, in eq 2, for $r > \mathcal{R}_{\text{ML}}$

$$\Delta E_{\alpha,r}(\bar{\mathbf{x}}) \rightarrow \Delta E_{\alpha,r}^{\text{ML}}(\bar{\mathbf{x}}) \quad (8)$$

and thus

$$E_{\text{sys,ML}}^{\text{graph-theoretic}}(\bar{\mathbf{x}}) = E^{\text{Ref}}(\bar{\mathbf{x}}) + \sum_{r=0}^{\mathcal{R}_{\text{ML}}} \sum_{\alpha \in \mathcal{V}_r} [\mathcal{M}_{\alpha,r}^{\mathcal{R}}] \Delta E_{\alpha,r}(\bar{\mathbf{x}}) + \sum_{r=\mathcal{R}_{\text{ML}}+1}^{\mathcal{R}} \sum_{\alpha \in \mathcal{V}_r} [\mathcal{M}_{\alpha,r}^{\mathcal{R}}] \Delta E_{\alpha,r}^{\text{ML}}(\bar{\mathbf{x}}) \quad (9)$$

where the larger rank terms in eq 2 have been replaced with ML estimates. Thus, instead of directly training on the full system potential energies, we prepare a mutually independent family of neural networks for different types of fragments to obtain ML-models for $\Delta E_{\alpha,r}^{\text{ML}}(\bar{\mathbf{x}})$. The process is illustrated in Figure 4. This approach has the advantage that as system size grows, the feature space needed to extrapolate the energy does not grow, since the features are based on fragment sizes. For example, as seen in Figure 4a, when the full system is used to compute a machine learning model, the number of nodes in the hidden layers is extremely large (1704×6 for the case of the PFOA-water periodic system treated here, where the number of terms in the input layer is 852×3). Consequently, the number of terms in the single neural network that is created has $852 \times 3 \times 1704 \times 6 + 3 \times (1704 \times 6)^2 + 1704 \times 6 = 300$ million terms! The resultant optimization problem is hard because one needs to find a single solution to this 300 million-dimensional nonlinear problem.

By contrast, eq 3 produces a family of independent neural networks, and the number of terms in these neural networks can be seen from a visual inspection of Figure 4b to be orders of magnitude smaller than that in Figure 4a. When 3-body terms ($\mathcal{R}=2$) are included, the number of terms in the neural network corresponding to each node (bottom row network of Figure 4b) is $[1 \times 3 \times 2 \times 6 + 3 \times (2 \times 6)^2 + 12]$ which is roughly 500. For each independent edge the number of terms in one neural network is $[5 \times 3 \times 10 \times 6 + 3 \times (10 \times 6)^2 + 10 \times 6]$ which is on the order of 10,000, thus, between nodes and edges, there is a substantial reduction in complexity of the

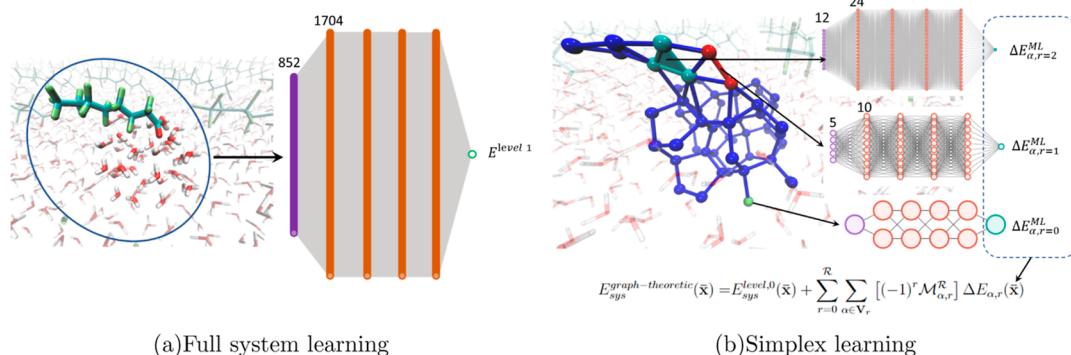


Figure 4. Comparison between (a) direct full system learning and (b) the machine learning process following graph-theoretic fragmentation. Here, we assume each graphic node contains 3 atoms. Every purple circle in the input layer of neural networks represents 3 features, and every orange circle in the hidden layer represents 6 neurons.

neural networks as compared to the full system, making the training process easier. For three-body interactions that are included in each independent face neural network, the number of parameters is $[12 \times 3 \times 24 \times 6 + 3 \times (24 \times 6)^2 + 24 \times 6]$ which is 67,536. The highest complexity here clearly arises from the face neural networks, but this is nearly 3 orders of magnitude lower in complexity with respect to the neural network for the full system. This can also graphically be seen from the density of neural network nodes in Figure 4a as compared to that in Figure 4b.

To achieve this, we start by constructing a data bank for all types of fragment geometries and energies that need to be replaced by machine learning predictions. Different types of fragments are separated for independent neural network models to learn their individual energy patterns. In order to capture the most representative geometries and reduce the cost of training, we apply a sampling method, Mini-batch- k -means,⁹⁰ based on some descriptors of the fragment geometries. Mini-batch- k -means is an efficient clustering algorithm that essentially tessellates a high dimensional space into a given number of components. These components are represented by some data inside, in our case, some typical fragment geometries. These geometries and their corresponding energies are then used in neural networks to find their energy patterns. Once all models are properly trained, they can efficiently produce accurate predictions as $\Delta E_{\alpha,r}^{\text{ML}}(\bar{\mathbf{x}})$ in eq 10 to accelerate the process of computing full system potential energies. More details about the descriptor of the geometries and Mini-batch- k -means can be found in Appendix A.

It is critical to note that for the applications considered in this paper, we benefit from efficient periodic DFT implementations¹⁷¹ that allow us to transfer the most critical computational bottlenecks to the larger fragment post-Hartree–Fock calculations. However, it is foreseeable that for extremely large unit cells, one may also need to use an ML approximation to the periodic DFT energy thus resulting in an effective graph-theory attenuated ML energy expression

$$E_{\text{sys,ML}}^{\text{graph-theoretic}}(\bar{\mathbf{x}}) = E^{\text{Ref;ML}}(\bar{\mathbf{x}}) + \sum_{r=0}^{\mathcal{R}_{\text{ML}}} \sum_{\alpha \in \mathbf{V}_r} [\mathcal{M}_{\alpha,r}^{\mathcal{R}}] \Delta E_{\alpha,r}(\bar{\mathbf{x}}) + \sum_{r=\mathcal{R}_{\text{ML}}+1}^{\mathcal{R}} \sum_{\alpha \in \mathbf{V}_r} [\mathcal{M}_{\alpha,r}^{\mathcal{R}}] \Delta E_{\alpha,r}^{\text{ML}}(\bar{\mathbf{x}}) \quad (10)$$

where both the periodic DFT as well as larger system post-Hartree–Fock energies are computed from ML.

3. INCLUDING DISPERSION CORRECTIONS WITHIN PERIODIC DFT THROUGH GRAPH-THEORETIC FRAGMENTATION AUGMENTED BY MACHINE LEARNING METHODS

We consider water surface interactions with PFOA where nonbonded vdW's type interactions, and fluxional hydrogen bonds between the adsorbate and water, play a prominent role. To study the adsorption of fluorocarbons on the surface of water, the ideal choice of electronic structure would capture these weak interactions while remaining computationally tractable for AIMD and potential energy surface calculations. Although DFT methods have proven to be useful for a wide range of systems, they have well documented shortcomings, including sizable self-interaction errors and difficulty properly capturing higher order dispersion interactions.^{23–26,42,184} For

example, previous studies on dynamics of water^{185,186} indicate that the bulk properties recovered vary significantly by the choice of DFT functional. Other studies on the interactions of fluorocarbons with water and other small molecules^{187–189} indicates that there may be a need to consider post-Hartree–Fock quality treatment of intermolecular interactions involving such polarizable systems.

As the goal here is to, eventually, consider extended systems, we choose to employ periodic condensed phase methods^{21,30,190–193} to capture the bulk physics of the system. Unfortunately, the inherent costs of this class of problems are significantly increased with the inclusion of nonlocal exchange^{29–35} and correlation,^{194–199} thereby limiting the accuracy of $E^{\text{Ref}}(\bar{\mathbf{x}})$ in eqs 2 and 10 to standard treatments of periodic electronic structure within semilocal DFT methods such as PBE²⁰⁰ and BLYP.²⁰¹ This limitation becomes even more pronounced when considering ab initio molecular dynamics simulations.^{22,35} In order to model the absorption of fluorocarbons, such as PFOA, on the surface of water, nonbonded interactions between water and fluorocarbons would need to be properly captured.

3.1. AIMD Trajectories Using eq 2: Generating the Test Set for Gauging Accuracy and Efficiency. To gauge the effectiveness of DFT, with and without empirical dispersion corrections, ab initio molecular dynamics trajectories were computed for PFOA on the surface of water in the condensed phase. Figure 1 shows an example unit cell with PFOA interacting with the surface of water. These trajectories were computed in the fashion discussed in ref 50 using eq 2 to extrapolate from PBE to CCSD, where the gradients of energy are written as

$$\frac{\partial E_{\text{PBC}}^{\text{graph-theoretic}}(\bar{\mathbf{x}})}{\partial \bar{\mathbf{x}}} = \frac{\partial E^{\text{Ref}}(\bar{\mathbf{x}})}{\partial \bar{\mathbf{x}}} + \sum_{r=0}^{\mathcal{R}} \sum_{\alpha \in \mathbf{V}_r} [\mathcal{M}_{\alpha,r}^{\mathcal{R}}] \frac{\partial \Delta E_{\alpha,r}(\bar{\mathbf{x}})}{\partial \bar{\mathbf{x}}_{\alpha,r}} \frac{\partial \bar{\mathbf{x}}_{\alpha,r}}{\partial \bar{\mathbf{x}}} \quad (11)$$

If bonds are broken during the formation of nodes in a graph, link atoms are used and hence $\bar{\mathbf{x}}_{\alpha,r}$ may not entirely be a subset of $\bar{\mathbf{x}}$. Thus, the term $\frac{\partial \bar{\mathbf{x}}_{\alpha,r}}{\partial \bar{\mathbf{x}}}$ in eq 11 is a Jacobian term which transforms the rank- r fragment gradient back to the full system gradient. The periodic full system calculation is performed with Quantum Espresso²⁰² with a kinetic energy cutoff of 50 Ry and Kresse-Joubert style pseudopotential.¹⁶¹ The resultant energy and forces are then augmented by short-range corrections arising from the graph theoretic representation given in eq 2. The fragments obtained from such a graphical decomposition were processed using Gaussian electronic structure suite¹⁶⁸ with the 6-31+G(d) basis. Thus, extending the standard notations used to represent composite treatments^{139,152} our periodic AIMD calculation may be characterized as, CCSD/6-31+G(d):PBE/6-31+G(d):PBE/50Ryd. The graphical representation used a $\mathcal{R}$ value of 2, which includes nodes, edges, and triangles with a distance cutoff of 4.0 Å for organic nodes and 3.0 Å for water nodes. These calculations are equilibrated to about 350 K within a microcanonical ensemble. Structures were sampled from the AIMD trajectory, as shown in Figure 5. These structures are sampled every 20 fs as representatives of the overall configurations encountered during the trajectory. In each of these configurations, we extracted the PFOA and the 4 directly

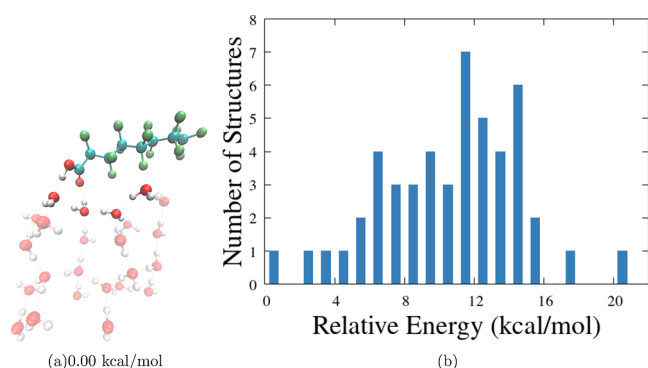


Figure 5. Geometries of PFOA and the four closest water molecules were sampled at 20 fs increments from the PFOA·(H₂O)₂₄ AIMD trajectory. The four closest waters in the initial structure are illustrated in (a), and these water molecules directly interact with the PFOA adsorbate. Furthermore, choosing a subsystem with four water molecules with the adsorbate provides a cluster size that is computationally feasible to handle with moderate basis CCSD, as needed in our computational benchmarks. The energy distribution of each of the 49 structures sampled from AIMD is illustrated in (b) at CCSD/aug-pVDZ.

interacting water molecules, as can be seen highlighted in Figure 5a. This provides a computationally tenable system to probe the quality of the different electronic structure methods and to gauge the practical effectiveness of eq 2 to capture many-body correlation effects at a much reduced computational time.

3.2. Accuracy of eq 2: the Need for Three-Body Interaction Terms. The chosen set of structures has a range of approximately 20 kcal/mol, as shown in Figure 5b. From each of these periodic structures, a cluster is carved out as noted above, which includes one PFOA molecule and four nearest neighbor water molecules. This is done to benchmark our graph theoretic expressions with full system CCSD results, and clearly the periodic system is much too large and impossible to compute at the CCSD level of theory. These structures of PFOA with 4 adjacent waters are then treated at the CCSD level to provide a background for benchmark to accurately account for long-range dispersion interactions. The conformational energies of these structures are calculated using aug-cc-pVDZ basis, as large basis sets are known to be required

to compute weak interactions accurately. The conformational energies were evaluated for a number of methods and basis sets and compared against CCSD/aug-cc-pVDZ. We discuss only the improvements one can make to the PBE functional by using eq 2. What is implied here is that $E^{\text{Ref}}(\bar{\mathbf{x}})$ in eq 2 and the perturbative (second) term can be viewed as providing improvements to the PBE level. Our results are summarized here in Figures 6–8.

As can be seen from Figure 6, the choice of basis sets appears to have a significant effect on agreement with the target energies as a choice of methods. It is clear that the semilocal functional PBE is insufficient to recover the necessary physical interactions for such weak interactions. Hybrid DFT approaches, such as PBE0 and B3LYP, in Figure 6 performed significantly better than the semilocal approaches. Yet considering that the fundamental goal here is the treatment of adsorbates on a surface, better performance of the nonlocal PBE0 or B3LYP does not satisfy the underlying purpose of this study, due to cost of hybrid functionals for periodic systems. Previous studies have also suggested that Halogen noncovalent interactions are challenging and the interactions there are only captured properly by high quality wave function theories.^{187–189,203,204} In order to overcome this obstacle, we present our graph-theoretic approach to embed dispersion into these calculations without the necessity of treating the whole system with the higher order scaling inherent to post-Hartree–Fock approaches.^{194–199,205} This graph-theoretic method relies upon a graphical representation of the system which has two principle parameters: edge cutoff and maximum rank, $\mathcal{R}$ (see eq 2). Figure 7a demonstrates the quality of this approach while varying both the edge cutoff, which governs the distance of the locality of the correlation correction, and the maximum rank, which controls the order of this correction. It is clear that corrections utilizing edges ($\mathcal{R} = 1$) offer remarkable improvement over the standard DFT calculations, but the three-body terms are necessary to obtain sub kcal/mol accuracy. (Also see Figure 7b.) The choice of an edge cutoff at 5.0 Å appears to capture the range of correlation-based weak interactions lacking in the underlying PBE functional with or without the empirical dispersion correction.

Next, using the graphical representation of $\mathcal{R} = 2$ and edge cutoff 5.0 Å, as informed by Figure 7, eq 2 is used to compare the structural ordering obtained using the larger basis, aug-cc-

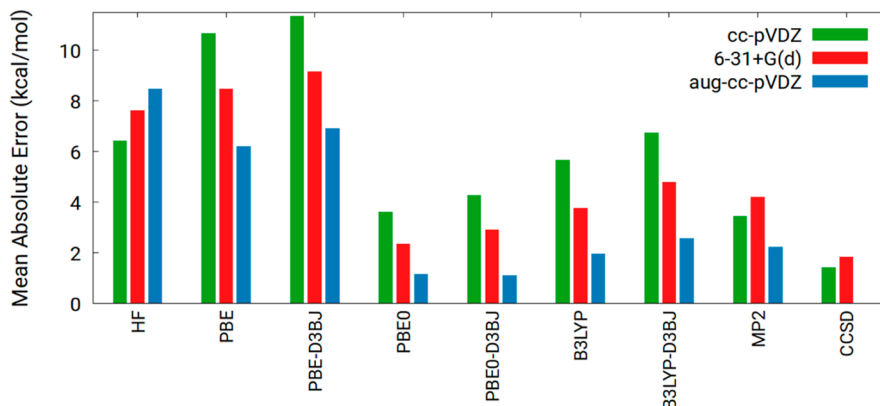


Figure 6. Structures discussed in Figure 5 were treated with a number of different electronic structure methods using three different basis sets, as shown with differing colored bars. The mean absolute error for this array of functionals and basis sets were computed in comparison to the target theory: CCSD/aug-cc-pVDZ. Large errors were observed for the semilocal functionals. No significant improvement was observed with the addition of empirical dispersion corrections to DFT functionals. These extrapolations demonstrate high accuracy with their full CCSD equivalents.

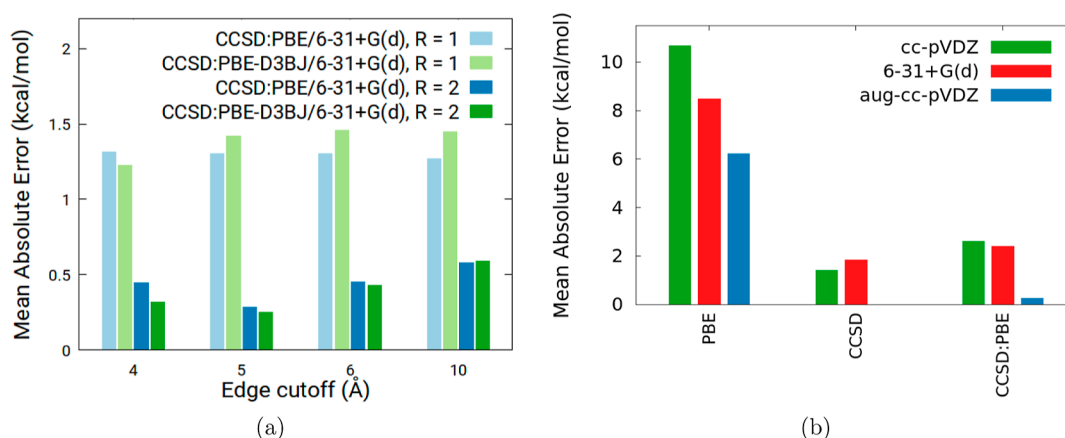


Figure 7. Geometries of PFOA with four water molecules from Figure 5 were calculated at CCSD/6-31+G(d) and their energies were extrapolated with a range of graphical representations. The quality of the extrapolations is compared against the energies from CCSD. From Figure (a) is clear that three body interactions ($R = 2$) are needed in the graphical representation to properly describe the water–fluorocarbon interactions and the fluorine–fluorine interactions in the fluorocarbon chain, and an edge cutoff of 5 Å is sufficient to capture the prerequisite interactions. In Figure (b), the PBE and CCSD results from Figure 6 are compared with the respective results for CCSD/PBE using the graphical representation using $R = 2$ and an edge cutoff of 5.0 Å. This graphical approach reproduces the CCSD results and massively improves upon the reference PBE results.

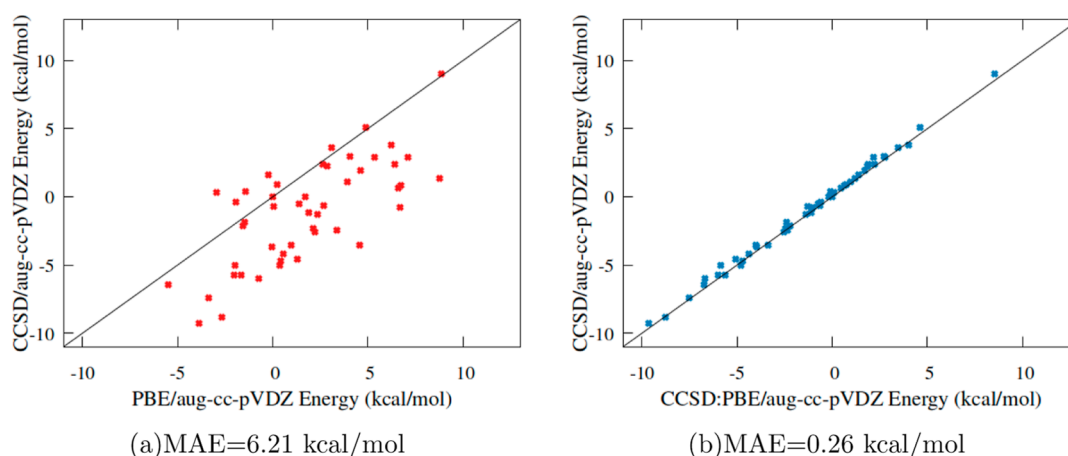


Figure 8. Figures (a,b) demonstrate the correlation of relative energies for semilocal DFT and the graph theoretic method using eq 2 against large basis CCSD quality energies. The graphical representation used a cutoff of 5 Å and $R = 2$. The graph-theoretic approach greatly improved the incorporation of dispersion and other weak interactions not captured within the PBE. The overall MAE for this set of structures is also noted.

Table 1. Neural Network Model Accuracy Using 5% Data as Training Set

triangle	total geometries	training set	MAE (kcal/mol)
H ₂ O-CF ₂ H ₂ -CO ₂ H ₂	1921	96	0.121
H ₂ O-CF ₂ H ₂ -CF ₂ H ₂	2585	129	0.180
(HCF ₂)(CF ₂)(CO ₂ H)	5843	292	0.051
(HCF ₂)(CF ₂)(CF ₂ H)	23364	1168	0.028
(HCF ₂)(CF ₂)(CF ₃)	5841	292	0.037

Table 2. Neural Network Model Accuracy Using 10% Data as Training Set

triangle	total geometries	training set	MAE (kcal/mol)
H ₂ O-CF ₂ H ₂ -CO ₂ H ₂	1921	192	0.047
H ₂ O-CF ₂ H ₂ -CF ₂ H ₂	2585	258	0.029
(HCF ₂)(CF ₂)(CO ₂ H)	5843	584	0.023
(HCF ₂)(CF ₂)(CF ₂ H)	23364	2336	0.005
(HCF ₂)(CF ₂)(CF ₃)	5841	584	0.010

pVDZ, in Figure 8b. These results are compared to that of PBE, the baseline treatment, Figure 8a, from which we correct

in Figure 8b. The overall errors to the graphical correlation correction are in the kilo-joule per mol (≈ 0.25 kcal/mol) range, while the PBE results have an MAE of over 6 kcal/mol error. These significantly improved energies give great promise to the ability to model weak interactions in the condensed phase accurately using the graph theoretical description.

3.3. Neural networks to Improve the Computational Efficiency of eq 2 through eq 10. The calculations above significantly reduce the computational complexity. The CPU time for a PFOA water system reduces from around 3 days (CCSD/aug-ccpVDZ) to 40 min (CCSD/PBE/aug-ccpVDZ). The necessity to include three body interactions involving fluorocarbon nodes implies that this approach is still limited by CCSD calculations on a sizable number of heavy atoms. Thus, we next explored the incorporation of transfer learning to further alleviate this new computational bottleneck.

Toward this, we first compute two condensed phase AIMD trajectories for a system containing one PFOA molecule on the surface of 24 and 48 water molecules. These simulations are used to produce sets of three body simplexes to be used as data sets for a transfer learning protocol discussed above and also in

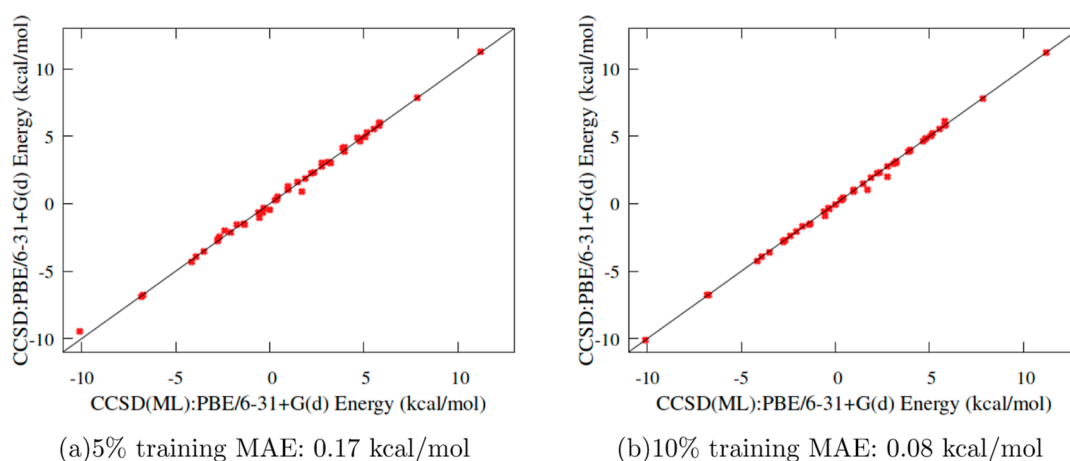


Figure 9. This figure demonstrates the correlation of potential energy between the graph-theoretic conformational energies and these graphical representations where the expensive triangular terms are replaced by machine learned energies (see Tables 1 and 2 in the Appendix A). The graphical representation used $\mathcal{R} = 2$ and a cutoff of 4 Å. Figures (a) used 5% of the data, while figures (b) used 10% of the data for training. The overall MAE values for these sets of structures are shown beneath the figures.

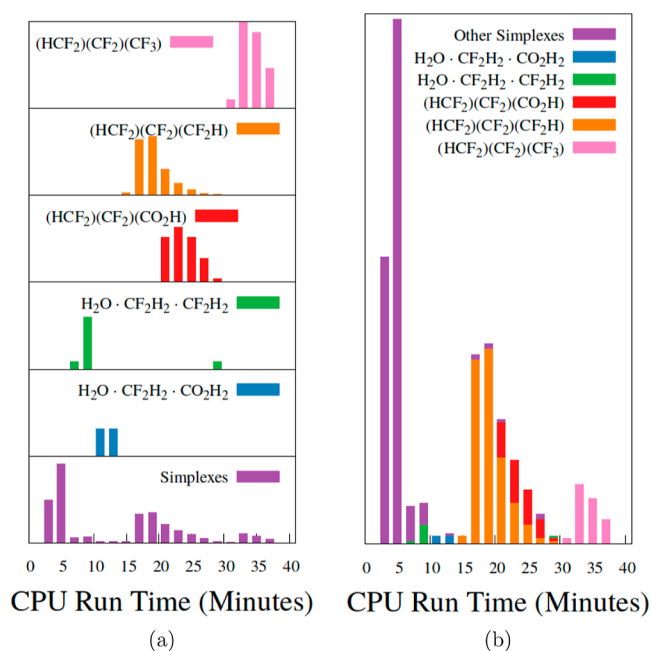


Figure A1. A detailed exploration of Figure 3. Run time costs for each simplex where the high cost simplexes are highlighted. These higher cost simplexes are then targets for our machine learning methodology. Figure (a) shows the distribution of the selected higher cost fragment terms that capture three body contributions with the overall distribution in the bottom panel. Figure (b) shows a stacked bar plot where the accumulative populations are shown binned together.

ref 70. Details regarding the method of selection of training data are described in Appendix A and follow the protocol established in ref 70. The neural networks set up involves two neural networks for each type of fragment (see eq 10) with the number of neurons defined, as in ref 70, as 4 times the number of input features (see Appendix A) in every hidden layer with 4 hidden layers to fit each type of rank 2 simplexes. Both neural networks use the same input vectors but the second network learns from the errors produced by the first.^{70,206} The individual neural network's accuracy is expressed in Tables 1 and 2 by using 5 and 10% of total data as training sets, respectively. These training sets are obtained from the Mini-

batch- k -means clustering algorithm discussed in Appendix A and in more detail in ref 89 and 207. After the training is completed, we evaluate the accuracy of all data points and replace the corresponding simplex energy term $\Delta E_{a,r}$ in eq 2 by the neural network predicted energy $\Delta E_{a,r}^{\text{ML}}$ (see eq 10) to compute the graph-theoretic full system potential energy. These models are then used to predict fragment energies from the AIMD trajectories of PFOA with 4 water molecules to compute the graph-theoretic potential energy at the edge cutoff of 4 Å. The results are displayed in Figure 9, where we observe an excellent transfer learning ability from smaller clusters to the large system with a full system MAE of 0.17 kcal/mol for the structural library in the previous section.

4. CONCLUSIONS

Semilocal DFT methods are commonly used for condensed phase electronic structure calculations, but these often struggle to properly capture weak interactions. While empirical dispersion corrections^{42,44} attempt to overcome this weakness, there are several new DFT functionals being developed²⁹ to address this issue, interactions in complex heterogeneous are challenging and corrections are often inadequate. This challenge was addressed in this work by employing our graph theoretic method to embed high quality correlation effects by capturing short-range, high order weak interactions between a fluorocarbon of environmental interest and a surface of water. Quality treatment of these surfaces was found to require three-body interaction terms to properly capture the target higher order correlation-based interactions, surpassing MP2 treatment of these interactions, and matching CCSD at the cost of semilocal DFT. Although this approach greatly increased the computational efficiency of obtaining high quality correlation treatments for these interactions, these calculations are also significantly impacted by the need for these, computationally intensive, higher order interaction terms inherent to the graphical fragmentation procedure. As a result, we also demonstrate the use of a recently introduced transfer machine learning procedure to create a family of Δ -machine learning models for different fragments to replicate the most expensive post-Hartree–Fock embedded cluster calculations from this graph-theoretic approach. This transfer learning approach shows minimal loss of accuracy in the

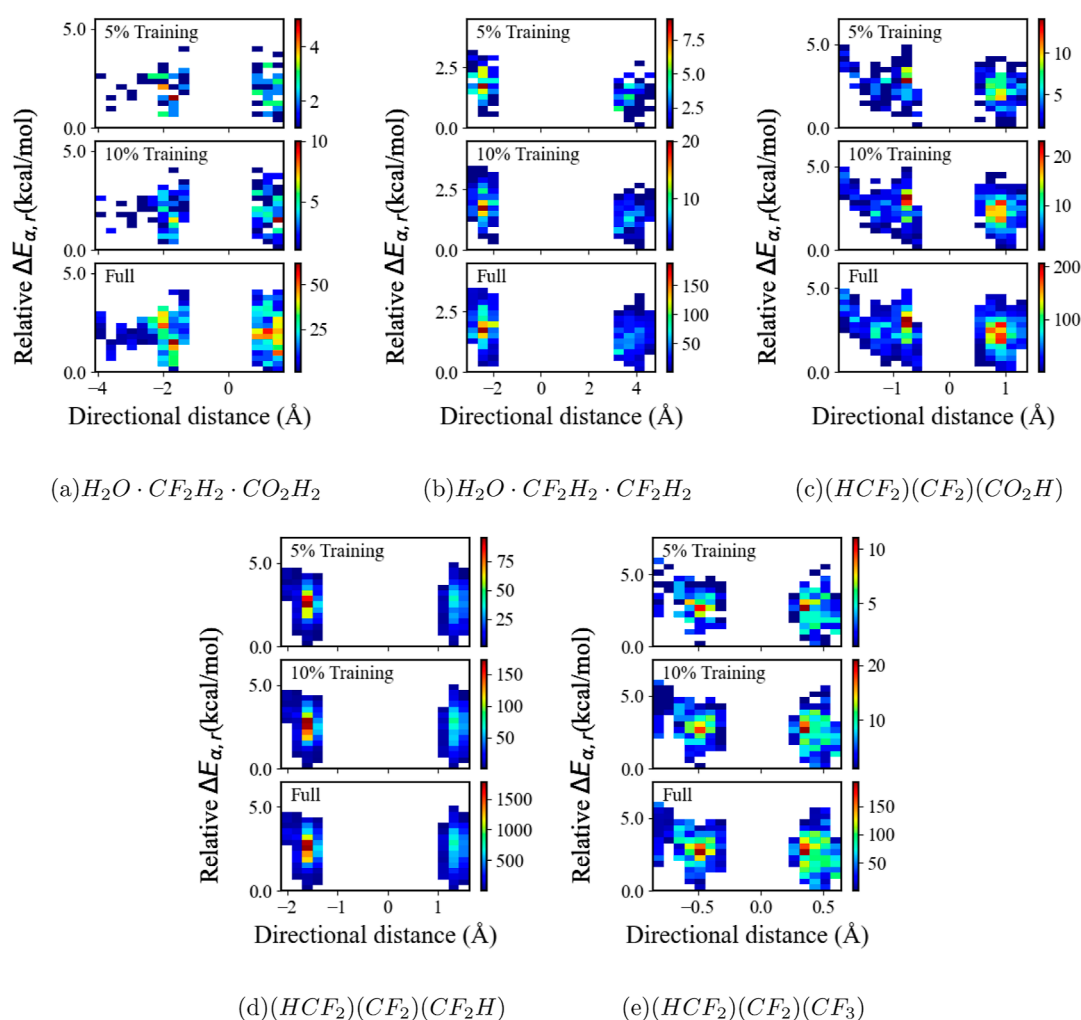


Figure A2. Training set distributions of 5 and 10% of data obtained from the Mini-batch-*k*-means algorithm, and all data points on the bottom panels labeled as “Full”.

graphical study and hence appears to successfully attain the coupled cluster accuracy for the target systems.

APPENDIX A

Transfer Machine Learning Protocol

Figure A1 shows the distribution of run time costs for all simplexes for the 50 structures discussed in the paper. The five most expensive fragments that capture three-body interactions represent a major bottleneck and are shown in Figure A1. These fragments are treated in Section 3.3 of the main paper using our transfer learning methodology. Details regarding the training of the machine learning model on these species are discussed below.

We follow the protocol established in ref 70. The vector $\mathbf{r}_i$ includes all interatomic distances for the *i*-th geometry of a given fragment. This is obtained from the distance matrix or Cartesian coordinates after these are sorted based on increasing atomic mass as done in ref 70. We then sample geometries from this set. The sampling method used is called mini-batch-*k*-means,^{89,207} which is a variant of the well-known *K*-means clustering algorithm;^{89,207} this approach tessellates or divides the data space for a given fragment in a geometric fashion into *k* mutually exclusive regions called clusters. Each cluster has a centroid $\bar{\mathbf{r}}_j$ and all the data points for a given

fragment, $\{\mathbf{r}_i\}$, are assigned to the closest centroid to create sets, C_j for each $\bar{\mathbf{r}}_j$. The *K*-means algorithm aims to find a preset number of centroids or clusters iteratively to minimize the cost function

$$\sum_{j=1}^k \sum_{\mathbf{r}_i \in C_j} |\mathbf{r}_i - \bar{\mathbf{r}}_j|^2 \quad (\text{A1})$$

During each iteration, all centroid positions are updated until no further change is seen for the position of centroids. The Mini-batch-*k*-means algorithm performs this task in an efficient way where only a random subset of data (known as a batch) is used to update the centroids during each iteration. After finding the converged centroid geometries, we use the set of closest data points to every centroid and form the training set that is then used to create the neural networks for each fragment.

In Figure A2, we display the training set distributions obtained from the above procedure as a 3-D histogram showing the relative density as different colors on the relative $\Delta E_{\alpha,r}$ vs directional distance panels. The $\Delta E_{\alpha,r}$ is a shifted energy difference between CCSD and DFT level of energy for each type of fragments. The directional distance serves as a reduced dimensional metric to measure the range of input vectors $\mathbf{r}_i$ and is defined as

$$\text{sign}[(\mathbf{r}_{\max} - \langle \mathbf{r} \rangle) \cdot (\mathbf{r}_i - \langle \mathbf{r} \rangle)] \times |\mathbf{r}_i - \langle \mathbf{r} \rangle| \quad (\text{A2})$$

where the $\text{sign}[\dots]$ represents the “sign” of the term within the parentheses. In, $\langle \mathbf{r} \rangle$ is an average geometry computed from a total number of geometries for a certain kind of fragment

$$\langle \mathbf{r} \rangle = \frac{1}{N} \sum_i \mathbf{r}_i \quad (\text{A3})$$

and a maximum geometry $\mathbf{r}_{\max}$ is used to define the positive direction

$$\mathbf{r}_{\max} = \arg \max_{\mathbf{r}_i} \|\mathbf{r}_i\|_2 \quad (\text{A4})$$

While the bottom panels in Figure A2 show the distributions of all data obtained from the two AIMD trajectories of PFOA with 24 and 48 water molecules inside a unit cell, the middle panels and top panels display the distributions when 10 and 5% of the fragments obtained from this AIMD data are used for training. Such data points are obtained from the above-mentioned sampling procedure, respectively. From Figure A2, we can see that the data distribution and relative density can both be well preserved even for a training ratio of only 5%, which helps to greatly reduce the number of computational intensive electronic structure calculations needed for the large fragments, and also avoid overfitting, which is a common problem in machine learning.

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

Authors

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

Xiao Zhu – Department of Chemistry and Department of Physics, Indiana University, Bloomington, Indiana 47405, United States

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.3c00955>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation grant CHE-2102610 to SSI. The computational facilities at Indiana University are duly acknowledged and have been critical to this effort. Acknowledgment is due in part to the Lilly Endowment, Inc., for their support of the BigRed computing facility at Indiana University widely used in the effort represented in this publication. This work was also supported in part by Shared University Research grants from IBM, Inc., to Indiana University, which supports the Scholarly Data Archives.

REFERENCES

- (1) Breslow, R.; Maitra, U.; Rideout, D. Selective Diels-Alder Reactions In Aqueous-Solutions And Suspensions. *Tetrahedron Lett.* **1983**, *24*, 1901–1904.
- (2) Breslow, R.; Maitra, U. On The Origin Of Product Selectivity In Aqueous Diels-Alder Reactions. *Tetrahedron Lett.* **1984**, *25*, 1239–1240.
- (3) Breslow, R. Hydrophobic Effects On Simple Organic-Reactions In Water. *Acc. Chem. Res.* **1991**, *24*, 159–164.
- (4) Narayan, S.; Muldoon, J.; Finn, M.; Fokin, V.; Kolb, H.; Sharpless, K. On Water: Unique reactivity of organic compounds in aqueous suspension. *Angew. Chem.* **2005**, *44*, 3275–3279.
- (5) Chanda, A.; Fokin, V. V. Organic Synthesis “On Water”. *Chem. Rev.* **2009**, *109*, 725–748.
- (6) Shapiro, N.; Vigalok, A. Highly Efficient Organic Reactions “on Water”, “in Water”, and Both. *Angew. Chem.* **2008**, *47*, 2849–2852.
- (7) Huang, J.; Zhang, X.; Armstrong, D. W. Highly Efficient Asymmetric Direct Stoichiometric Aldol Reactions On/In Water. *Angew. Chem.* **2007**, *46*, 9073–9077.
- (8) Turner, G. L.; Morris, J. A.; Greaney, M. F. Direct Arylation Of Thiazoles On Water. *Angew. Chem.* **2007**, *46*, 7996–8000.
- (9) Wu, X.; Li, X.; Zanotti-Gerosa, A.; Pettman, A.; Liu, J.; Mills, A. J.; Xiao, J. Rh(III)- And Ir(III)-Catalyzed Asymmetric Transfer Hydrogenation Of Ketones In Water. *Chem.—Eur. J.* **2008**, *14*, 2209–2222.
- (10) Butler, R. N.; Coyne, A. G.; Cunningham, W. J.; Moloney, E. M. Water And Organic Synthesis: A Focus On The In-Water And On-Water Border. Reversal Of The In-Water Breslow Hydrophobic Enhancement Of The Normal Endo-Effect On Crossing To On-Water Conditions For Huisgen Cycloadditions With Increasingly Insoluble Organic Liquid And Solid 2 Pi-Dipolarophiles. *J. Org. Chem.* **2013**, *78*, 3276–3291.
- (11) 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*, 796–802.
- (12) 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*, 3562–3565.
- (13) Jung, Y.; Marcus, R. A. On The Theory Of Organic Catalysis On Water. *J. Am. Chem. Soc.* **2007**, *129*, 5492–5502.
- (14) Jung, Y.; Marcus, R. A. Protruding Interfacial OH Groups And “On-Water” Heterogeneous Catalysis. *J. Phys.: Condens. Matter* **2010**, *22*, 284117.
- (15) Kühne, T. D.; Pascal, T. A.; Kaxiras, E.; Jung, Y. New Insights Into The Structure Of The Vapor/Water Interface From Large-Scale First-Principles Simulations. *J. Phys. Chem. Lett.* **2011**, *2*, 105–113.
- (16) 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.
- (17) Kessler, J.; Elgabarty, H.; Spura, T.; Karhan, K.; Partovi-Azar, P.; Hassanali, A. A.; Kühne, T. D. Structure And Dynamics Of The Instantaneous Water/Vapor Interface Revisited By Path-Integral And Ab Initio Molecular Dynamics Simulations. *J. Phys. Chem. B* **2015**, *119*, 10079–10086.
- (18) 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.* **2014**, *372*, 10120476.
- (19) 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*, 2315–2372.
- (20) Zhao, Y.; Truhlar, D. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (21) Kudin, K. N.; Scuseria, G. E.; Martin, R. L. Hybrid Density-Functional Theory and the Insulating Gap of UO₂. *Phys. Rev. Lett.* **2002**, *89*, 266402.
- (22) 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*, 141.

(23) Perdew, J.; Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1981**, *23*, 5048–5079.

(24) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289–320.

(25) Mori-Sánchez, 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.

(26) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into current limitations of density functional theory. *Science* **2008**, *321*, 792–794.

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

(28) Ohto, T.; Dodia, M.; Xu, J.; Imoto, S.; Tang, F.; Zysk, F.; Kühne, T. D.; Shigeta, Y.; Bonn, M.; Wu, X.; Nagata, Y. Accessing the Accuracy of Density Functional Theory through Structure and Dynamics of the Water–Air Interface. *J. Phys. Chem. Lett.* **2019**, *10*, 4914–4919.

(29) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

(30) Kudin, K. N.; Scuseria, G. E. Range definitions for Gaussian-type charge distributions in fast multipole methods. *J. Chem. Phys.* **1999**, *111*, 2351–2356.

(31) 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*, 443–454.

(32) Lucero, M.; Henderson, T.; Scuseria, G. Improved semiconductor lattice parameters and band gaps from a middle-range screened hybrid exchange functional. *J. Phys.: Condens. Matter* **2012**, *24*, 145504.

(33) 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.

(34) Bylaska, E. J. Plane-Wave DFT Methods for Chemistry. *Annual Reports in Computational Chemistry*; Elsevier, 2017; Vol. 13, pp 185–228.

(35) 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.

(36) Müller, C.; Usvyat, D. Incrementally Corrected Periodic Local MP2 Calculations: I. The Cohesive Energy of Molecular Crystals. *J. Chem. Theory Comput.* **2013**, *9*, 5590–5598.

(37) Pisani, C.; Schütz, M.; Casassa, S.; Usvyat, D.; Maschio, L.; Lorenz, M.; Erba, A. Cryscor: a program for the post-Hartree–Fock treatment of periodic systems. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7615.

(38) Kuo, I.-F. W.; Mundy, C. J. An ab Initio Molecular Dynamics Study of the Aqueous Liquid–Vapor Interface. *Science* **2004**, *303*, 658–660.

(39) Li, X.; Iyengar, S. S. Quantum Wavepacket Ab Initio Molecular Dynamics for Extended Systems. *J. Phys. Chem. A* **2011**, *115*, 6269–6284.

(40) Ohto, T.; Dodia, M.; Xu, J.; Imoto, S.; Tang, F.; Zysk, F.; Kühne, T. D.; Shigeta, Y.; Bonn, M.; Wu, X.; Nagata, Y. Accessing the Accuracy of Density Functional Theory through Structure and Dynamics of the Water–Air Interface. *J. Phys. Chem. Lett.* **2019**, *10*, 4914–4919.

(41) Tran, F.; Stelzl, J.; Blaha, P. Rungs 1 to 4 of DFT Jacob's ladder: Extensive test on the lattice constant, bulk modulus, and cohesive energy of solids. *J. Chem. Phys.* **2016**, *144*, 204120.

(42) Klimeš, J.; Michaelides, A. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **2012**, *137*, 120901.

(43) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements. *J. Chem. Phys.* **2010**, *132*, 154104.

(44) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(45) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(46) Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. Higher-accuracy van der Waals density functional. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 081101.

(47) Dovesi, R.; Erba, A.; Orlando, R.; Zicovich-Wilson, C. M.; Civalieri, B.; Maschio, L.; Rérat, M.; Casassa, S.; Baima, J.; Salustro, S.; Kirtman, B. Quantum-mechanical condensed matter simulations with CRYSTAL. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, No. e1360.

(48) Ko, H.-Y.; Jia, J.; Santra, B.; Wu, X.; Car, R.; DiStasio, R. A., Jr. Enabling large-scale condensed-phase hybrid density functional theory based ab initio molecular dynamics. 1. theory, algorithm, and performance. *J. Chem. Theory Comput.* **2020**, *16*, 3757–3785.

(49) Stein, F.; Hutter, J. Double-hybrid density functionals for the condensed phase: Gradients, stress tensor, and auxiliary-density matrix method acceleration. *J. Chem. Phys.* **2022**, *156*, 074107.

(50) Ricard, T. C.; Iyengar, S. S. Efficient and Accurate Approach To Estimate Hybrid Functional and Large Basis-Set Contributions to Condensed-Phase Systems and Molecule–Surface Interactions. *J. Chem. Theory Comput.* **2020**, *16*, 4790–4812.

(51) Varandas, A. J.; Murrell, J. N. A many-body expansion of polyatomic potential energy surfaces: application to H_n systems. *Faraday Discuss. Chem. Soc.* **1977**, *62*, 92.

(52) Murrell, J.; Carter, S.; Farantos, S.; Huxley, P.; Varandas, A. *Molecular Potential Energy Functions*; Wiley: New York, 1984.

(53) Xantheas, S. S. Ab Initio Studies of Cyclic Water Clusters $(H_2O)_n$, $n = 1–6$. II. Analysis of Many-body Interactions. *J. Chem. Phys.* **1994**, *100*, 7523–7534.

(54) Xantheas, S. S. Ab Initio Studies of Cyclic Water Clusters $(H_2O)_n$, $n = 1–6$. III. Comparison of Density Functional with MP2 Results. *J. Chem. Phys.* **1995**, *102*, 4505–4517.

(55) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many-Body Expansion for Large Systems, with Applications to Water Clusters. *J. Chem. Theory Comput.* **2007**, *3*, 46–53.

(56) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many-Body Expansion for Simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1–6.

(57) 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.

(58) 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.

(59) Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schütz, M.; Chan, G. K.-L. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science* **2014**, *345*, 640–643.

(60) Li, J.; Iyengar, S. S. Ab initio Molecular Dynamics using Recursive, Spatially Separated, Overlapping Model Subsystems Mixed Within an ONIOM Based Fragmentation Energy Extrapolation Technique. *J. Chem. Theory Comput.* **2015**, *11*, 3978–3991.

(61) Li, J.; Haycraft, C.; Iyengar, S. S. Hybrid, Extended Lagrangian – Born–Oppenheimer Ab Initio Molecular Dynamics using Fragment-Based Electronic Structure. *J. Chem. Theory Comput.* **2016**, *12*, 2493–2508.

(62) Haycraft, C.; Li, J.; Iyengar, S. S. Efficient, “On-the-fly” Born–Oppenheimer and Car–Parrinello–type Dynamics with coupled cluster accuracy through Fragment Based Electronic Structure. *J. Chem. Theory Comput.* **2017**, *13*, 21887–21901.

(63) 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*, 2852–2866.

- (64) 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*, 5535–5552.
- (65) 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*, 5769–5786.
- (66) 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 ab initio molecular dynamics, and potential surfaces for quantum nuclear effects. *Int. J. Quantum Chem.* **2020**, *120*, No. e26244.
- (67) Kumar, A.; DeGregorio, N.; Iyengar, S. S. Graph-Theory-Based Molecular Fragmentation for Efficient and Accurate Potential Surface Calculations in Multiple Dimensions. *J. Chem. Theory Comput.* **2021**, *17*, 6671–6690.
- (68) Zhang, J. H.; Ricard, T. C.; Haycraft, C.; Iyengar, S. S. Weighted-Graph-Theoretic Methods for Many-Body Corrections within ONIOM: Smooth AIMD and the Role of High-Order Many-Body Terms. *J. Chem. Theory Comput.* **2021**, *17*, 2672–2690.
- (69) Zhang, J. H.; Iyengar, S. S. Graph-|Q>(Cl): A Graph-based Quantum-classical algorithm for efficient electronic structure on hybrid quantum/classical hardware systems: Improved quantum circuit depth performance. *J. Chem. Theory Comput.* **2022**, *18*, 2885–2899.
- (70) Zhu, X.; Iyengar, S. S. Graph Theoretic Molecular Fragmentation for Multidimensional Potential Energy Surfaces Yield an Adaptive and General Transfer Machine Learning Protocol. *J. Chem. Theory Comput.* **2022**, *18*, 5125–5144.
- (71) Kumar, A.; DeGregorio, N.; Ricard, T.; Iyengar, S. S. Graph-Theoretic Molecular Fragmentation for Potential Surfaces Leads Naturally to a Tensor Network Form and Allows Accurate and Efficient Quantum Nuclear Dynamics. *J. Chem. Theory Comput.* **2022**, *18*, 7243–7259.
- (72) Bishop, C. M. *Pattern Recognition and Machine Learning*; Springer: New York, 2006.
- (73) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohli, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
- (74) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710.
- (75) Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121*, 10037–10072.
- (76) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K. R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142–10186.
- (77) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- (78) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. Δ – machine learning for potential energy surfaces: A PIP approach to bring a DFT-based PES to CCSD(T) level of theory. *J. Chem. Phys.* **2021**, *154*, 051102.
- (79) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ -Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.
- (80) Chen, J.; Xu, W.; Zhang, R. Machine learning-driven discovery of double hybrid organic–inorganic perovskites. *J. Mater. Chem. A* **2022**, *10*, 1402–1413.
- (81) Unzueta, P. A.; Greenwell, C. S.; Beran, G. J. O. Predicting Density Functional Theory-Quality Nuclear Magnetic Resonance Chemical Shifts via -Machine Learning. *J. Chem. Theory Comput.* **2021**, *17*, 826–840.
- (82) Yoo, D.; Lee, K.; Jeong, W.; Lee, D.; Watanabe, S.; Han, S. Atomic energy mapping of neural network potential. *Phys. Rev. Mater.* **2019**, *3*, 093802.
- (83) Marques, M. R. G.; Wolff, J.; Steigemann, C.; Marques, M. A. L. Neural network force fields for simple metals and semiconductors: construction and application to the calculation of phonons and melting temperatures. *Phys. Chem. Chem. Phys.* **2019**, *21*, 6506–6516.
- (84) Liu, M.; Kitchin, J. R. SingleNN: Modified Behler–Parrinello Neural Network with Shared Weights for Atomistic Simulations with Transferability. *J. Phys. Chem. C* **2020**, *124* (32), 17811–17818.
- (85) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **2017**, *8*, 13890.
- (86) Yao, K.; Herr, J. E.; Parkhill, J. The many-body expansion combined with neural networks. *J. Chem. Phys.* **2017**, *146*, 146.
- (87) Chen, W.-K.; Fang, W.-H.; Cui, G. Integrating machine learning with the multilayer energy-based fragment method for excited states of large systems. *J. Phys. Chem. Lett.* **2019**, *10*, 7836–7841.
- (88) Cheng, Z.; Zhao, D.; Ma, J.; Li, W.; Li, S. An on-the-fly approach to construct generalized energy-based fragmentation machine learning force fields of complex systems. *J. Phys. Chem. A* **2020**, *124*, S007–S014.
- (89) Hartigan, J. A.; Wong, M. A. Algorithm AS 136: A k-means clustering algorithm. *J. Roy. Stat. Soc. C Appl. Stat.* **1979**, *28*, 100–108.
- (90) Sculley, D. Web-scale k-means clustering. *WWW’10: Proceedings of the 19th international conference on World wide web*; Association for Computing Machinery: New York, NY, USA, 2010, pp 1177–1178.
- (91) Knuth, D. E. *The Art of Computer Programming*; Addison-Wesley: Reading, MA, 1973.
- (92) Manber, U. *Introduction to Algorithms: A Creative Approach*; Pearson Education: NJ, 1989.
- (93) Faithfull, N.; Weers, J. Perfluorocarbon compounds. *Vox Sang.* **1998**, *74*, 243–248.
- (94) Lemal, D. Perspective on fluorocarbon chemistry. *J. Org. Chem.* **2004**, *69*, 1–11.
- (95) Kudo, N.; Kawashima, Y. Toxicity and toxicokinetics of perfluorooctanoic acid in humans and animals. *J. Toxicol. Sci.* **2003**, *28*, 49–57.
- (96) Lau, C.; Anitole, K.; Hodes, C.; Lai, D.; Pfahles-Hutchens, A.; Seed, J. Perfluoroalkyl acids: A review of monitoring and toxicological findings. *Toxicol. Sci.* **2007**, *99*, 366–394.
- (97) Jansen, A.; Müller, M. H.; Grønnestad, R.; Klungsoyr, O.; Polder, A.; Skjerve, E.; Aaseth, J.; Lyche, J. L. Decreased plasma levels of perfluoroalkylated substances one year after bariatric surgery. *Sci. Total Environ.* **2019**, *657*, 863–870.
- (98) Richterová, D.; Fáblová, L.; Patayová, H.; Pulkrabová, J.; Lanková, D.; Rausová, K.; Šovčíková, E.; Štencl, J.; Hajšlová, J.; Trnovec, T.; et al. Determinants of prenatal exposure to perfluoroalkyl substances in the Slovak birth cohort. *Environ. Int.* **2018**, *121*, 1304–1310.
- (99) Luft, C. M.; Schutt, T. C.; Shukla, M. K. Properties and Mechanisms for PFAS Adsorption to Aqueous Clay and Humic Soil Components. *Environ. Sci. Technol.* **2022**, *56*, 10053–10061.
- (100) Sundararaman, R.; Arias, T. Regularization of the Coulomb singularity in exact exchange by Wigner-Seitz truncated interactions: Towards chemical accuracy in nontrivial systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 165122.
- (101) Fromme, H.; Tittlemier, S. A.; Völkel, W.; Wilhelm, M.; Twardella, D. Perfluorinated compounds - Exposure assessment for

the general population in western countries. *Int. J. Hyg Environ. Health* **2009**, *212*, 239–270.

(102) Vestergren, R.; Cousins, I. T. Tracking the Pathways of Human Exposure to Perfluorocarboxylates. *Environ. Sci. Technol.* **2009**, *43*, 5565–5575.

(103) Haukås, M.; Berger, U.; Hop, H.; Gulliksen, B.; Gabrielsen, G. W. Bioaccumulation of per- and polyfluorinated alkyl substances (PFAS) in selected species from the Barents Sea food web. *Environ. Pollut.* **2007**, *148*, 360–371.

(104) Key, B. D.; Howell, R. D.; Criddle, C. S. Fluorinated Organics in the Biosphere. *Environ. Sci. Technol.* **1997**, *31*, 2445–2454.

(105) Washington, J. W.; Jenkins, T. M.; Rankin, K.; Naile, J. E. Decades-Scale Degradation of Commercial, Side-Chain, Fluorotelomer-Based Polymers in Soils and Water. *Environ. Sci. Technol.* **2015**, *49*, 915–923.

(106) Sunderland, E. M.; Hu, X. C.; Dassuncao, C.; Tokranov, A. K.; Wagner, C. C.; Allen, J. G. A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects. *J. Expo. Sci. Environ. Epidemiol.* **2019**, *29*, 131–147.

(107) Melzer, D.; Rice, N.; Depledge, M. H.; Henley, W. E.; Galloway, T. S. Association between serum perfluorooctanoic acid (PFOA) and thyroid disease in the US National Health and Nutrition Examination Survey. *Environ. Health Perspect.* **2010**, *118*, 686–692.

(108) Shankar, A.; Xiao, J.; Ducatman, A. Perfluoroalkyl chemicals and chronic kidney disease in US adults. *Am. J. Epidemiol.* **2011**, *174*, 893–900.

(109) Barry, V.; Winquist, A.; Steenland, K. Perfluorooctanoic acid (PFOA) exposures and incident cancers among adults living near a chemical plant. *Environ. Health Perspect.* **2013**, *121*, 1313–1318.

(110) Steenland, K.; Fletcher, T.; Savitz, D. A. Epidemiologic Evidence on the Health Effects of Perfluorooctanoic Acid (PFOA). *Environ. Health Perspect.* **2010**, *118*, 1100–1108.

(111) Steenland, K.; Woskie, S. Cohort mortality study of workers exposed to perfluorooctanoic acid. *Am. J. Epidemiol.* **2012**, *176*, 909–917.

(112) Fei, C.; McLaughlin, J. K.; Tarone, R. E.; Olsen, J. Perfluorinated chemicals and fetal growth: a study within the Danish National Birth Cohort. *Environ. Health Perspect.* **2007**, *115*, 1677–1682.

(113) Joensen, U. N.; Bossi, R.; Leffers, H.; Jensen, A. A.; Skakkebaek, N. E.; Jorgensen, N. Do perfluoroalkyl compounds impair human semen quality? *Environ. Health Perspect.* **2009**, *117*, 923–927.

(114) Renner, R. PFOS phaseout pays off. *Environ. Sci. Technol.* **2008**, *42*, 4618.

(115) Lindstrom, A. B.; Strynar, M. J.; Libelo, E. L. Polyfluorinated Compounds: Past, Present, and Future. *Environ. Sci. Technol.* **2011**, *45*, 7954–7961.

(116) Post, G. B.; Cohn, P. D.; Cooper, K. R. Perfluorooctanoic acid (PFOA), an emerging drinking water contaminant: a critical review of recent literature. *Environ. Res.* **2012**, *116*, 93–117.

(117) USEPA. Announcement of Final Regulatory Determinations for Contaminants on the Fourth Drinking Water Contaminant Candidate List. *Federal Register*; USEPA, 2021; Vol. 86, p 12272.

(118) Evich, M. G.; Davis, M. J. B.; McCord, J. P.; Acrey, B.; Awkerman, J. A.; Knappe, D. R. U.; Lindstrom, A. B.; Speth, T. F.; Tebes-Stevens, C.; Strynar, M. J.; Wang, Z.; Weber, E. J.; Henderson, W. M.; Washington, J. W. Per- and polyfluoroalkyl substances in the environment. *Science* **2022**, *375*, No. eabg9065.

(119) Joudan, S.; Lundgren, R. J. Taking the “F” out of forever chemicals. *Science* **2022**, *377*, 816–817.

(120) Trang, B.; Li, Y.; Xue, X.-S.; Ateia, M.; Houk, K. N.; Dichtel, W. R. Low-temperature mineralization of perfluorocarboxylic acids. *Science* **2022**, *377*, 839–845.

(121) 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*, 1563–1568.

(122) Zhang, D. W.; Zhang, J. Z. H. Molecular Fractionation with Conjugate Caps for Full Quantum Mechanical Calculation of Protein–molecule Interaction Energy. *J. Chem. Phys.* **2003**, *119*, 3599–3605.

(123) Polkosnik, W.; Massa, L. Kohn-Sham Density Matrix and the Kernel Energy Method. *Acta Phys.-Chim. Sin.* **2018**, *34*, 656.

(124) Stoll, H.; Paulus, B.; Fulde, P. On the accuracy of correlation-energy expansions in terms of local increments. *J. Chem. Phys.* **2005**, *123*.

(125) Paulus, B. The method of increments - a wavefunction-based ab initio correlation method for solids. *Phys. Rep.* **2006**, *428*, 1–52.

(126) 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.

(127) 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.

(128) Kamiya, M.; Hirata, S.; Valiev, M. Fast electron correlation methods for molecular clusters without basis set superposition errors. *J. Chem. Phys.* **2008**, *128*, 074103.

(129) Guo, W.; Wu, A.; Xu, X. X. O. XO: An extended ONIOM method for accurate and efficient geometry optimization of large molecules. *Chem. Phys. Lett.* **2010**, *498*, 203–208.

(130) Mayhall, N. J.; Raghavachari, K. Molecules-In-Molecules: An Extrapolated Fragment-Based Approach for Accurate Calculations on Large Molecules and Materials. *J. Chem. Theory Comput.* **2011**, *7*, 1336–1343.

(131) Le, H.-A.; Tan, H.-J.; Ouyang, J. F.; Bettens, R. P. A. Combined Fragmentation Method: A Simple Method for Fragmentation of Large Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 469–478.

(132) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632–672.

(133) Mayhall, N. J.; Raghavachari, K. Many-Overlapping-Body (MOB) Expansion: A Generalized Many Body Expansion for Nondisjoint Monomers in Molecular Fragmentation Calculations of Covalent Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 2669–2675.

(134) Beran, G. J. O.; Hirata, S. Fragment and Localized Orbital Methods in Electronic Structure Theory. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7559.

(135) Collins, M. A. Systematic Fragmentation of Large Molecules by Annihilation. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7744.

(136) Sahu, N.; Yeole, S. D.; Gadre, S. R. Appraisal of molecular tailoring approach for large clusters. *J. Chem. Phys.* **2013**, *138*, 104101.

(137) Li, S.; Li, W.; Ma, J. Generalized Energy-Based Fragmentation Approach and Its Applications to Macromolecules and Molecular Aggregates. *Acc. Chem. Res.* **2014**, *47*, 2712–2720.

(138) 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; Springer, 2014; Vol. 345, p 59.

(139) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643–5677.

(140) Saha, A.; Raghavachari, K. Analysis of different fragmentation strategies on a variety of large peptides: Implementation of a low level of theory in fragment-based methods can be a crucial factor. *J. Chem. Theory Comput.* **2015**, *11*, 2012–2023.

(141) 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.

(142) Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.* **2015**, *115*, 5607–5642.

- (143) 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*, 708–713.
- (144) Beran, G. J. O. Modeling polymorphic molecular crystals with electronic structure theory. *Chem. Rev.* **2016**, *116*, 5567–5613.
- (145) 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*, 22695–22699.
- (146) 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*, 1413–1420.
- (147) 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*, 475–487.
- (148) Nyden, M.; Petersson, G. Complete basis set correlation energies. I. The asymptotic convergence of pair natural orbital expansions. *J. Chem. Phys.* **1981**, *75*, 1843–1862.
- (149) Pople, J.; Head-Gordon, M.; Fox, D.; Raghavachari, K.; Curtiss, L. Gaussian-1 theory: A general procedure for prediction of molecular energies. *J. Chem. Phys.* **1989**, *90*, 5622–5629.
- (150) Montgomery, J.; Frisch, M.; Ochterski, J.; Petersson, G. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J. Chem. Phys.* **1999**, *110*, 2822–2827.
- (151) DeYonker, N.; Cundari, T.; Wilson, A. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. *J. Chem. Phys.* **2006**, *124*, 114104.
- (152) Herbert, J. M. Fantasy versus reality in fragment-based quantum chemistry. *J. Chem. Phys.* **2019**, *151*, 170901.
- (153) 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.
- (154) 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.
- (155) 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*, 6799–6805.
- (156) Meitei, O. R.; Heßelmann, A. On the Stability of Cyclophane Derivates Using a Molecular Fragmentation Method. *ChemPhysChem* **2016**, *17*, 3863–3874.
- (157) 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*, 2200–2214.
- (158) Li, W.; Dong, H.; Ma, J.; Li, S. Structures and spectroscopic properties of large molecules and condensed-phase systems predicted by generalized energy-based fragmentation approach. *Acc. Chem. Res.* **2021**, *54*, 169–181.
- (159) Liu, J.; He, X. Recent advances in quantum fragmentation approaches to complex molecular and condensed-phase systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2023**, *13*, No. e1650.
- (160) Blochl, P. Projector Augmented-Wave Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953–17979.
- (161) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758–1775.
- (162) Svensson, M.; Humbel, S.; Froese, R. D.; 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 $\text{Pt}(\text{P}(\text{t-Bu})_3)_2 + \text{H}_2$ oxidative addition. *J. Phys. Chem.* **1996**, *100*, 19357–19363.
- (163) 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*, 327–350.
- (164) 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*, 5678–5796.
- (165) Nagayoshi, K.; Ikeda, T.; Kitaura, K.; Nagase, S. Computational Procedure of Lattice Energy Using the Ab Initio MO Method. *J. Theor. Comput. Chem.* **2003**, *02*, 233–244.
- (166) Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. *Chem. Rev.* **2016**, *116*, 7501–7528.
- (167) Yu, Q.; Bowman, J. M. Communication: VSCF/VCI vibrational spectroscopy of H_2O 3+ and H_2O 4+ using high-level, many-body potential energy surface and dipole moment surfaces. *J. Chem. Phys.* **2017**, *146*, 121102.
- (168) 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.
- (169) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 73–78.
- (170) 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. Chem. Theory Comput.* **2017**, *13*, 3185–3197.
- (171) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter.* **2009**, *21*, 395502.
- (172) Warshel, A.; Weiss, R. M. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.
- (173) Chang, Y. T.; Miller, W. H. An empirical valence bond model for constructing global potential energy surfaces for chemical reactions of polyatomic molecular systems. *J. Phys. Chem.* **1990**, *94*, 5884–5888.
- (174) Day, T. J.; Soudackov, A. V.; Čuma, M.; Schmitt, U. W.; Voth, G. A. A second generation multistate empirical valence bond model for proton transport in aqueous systems. *J. Chem. Phys.* **2002**, *117*, 5839–5849.
- (175) Baer, M. *Beyond Born-Oppenheimer: Conical Intersections and Electronic Nonadiabatic Coupling Terms*; Wiley: New York, 2006.
- (176) Hanna, G.; Kapral, R. Nonadiabatic Dynamics of Condensed Phase Rate Processes. *Acc. Chem. Res.* **2006**, *39*, 21–27.

- (177) Worth, G. A.; Cederbaum, L. S. Beyond Born-Oppenheimer: Molecular Dynamics Through a Conical Intersection. *Annu. Rev. Phys. Chem.* **2004**, *55*, 127–158.
- (178) Jasper, A. W.; Zhu, C.; Nangia, S.; Truhlar, D. G. Introductory Lecture: Nonadiabatic Effects in Chemical Dynamics. *Faraday Discuss.* **2004**, *127*, 1.
- (179) Kendrick, B. K.; Alden Mead, C.; Truhlar, D. G. Properties of Nonadiabatic Couplings and the Generalized Born Oppenheimer Approximation. *Chem. Phys.* **2002**, *277*, 31–41.
- (180) Kuppermann, A. The Geometric Phase in Reaction Dynamics. In *Dynamics of Molecules and Chemical Reactions*; Wyatt, R. E., Zhang, J. Z. H., Eds.; Marcel Dekker Inc.: New York, NY, 1996; p 411.
- (181) Yarkony, D. R. Diabolical Conical Intersections. *Rev. Mod. Phys.* **1996**, *68*, 985–1013.
- (182) Matsika, S.; Yarkony, D. R. Beyond Two-State Conical Intersections. Three-State Conical Intersections in Low Symmetry Molecules: The Allyl Radical. *J. Am. Chem. Soc.* **2003**, *125*, 10672–10676.
- (183) Lin, H.; Zhao, Y.; Tishchenko, O.; Truhlar, D. G. Multiconfiguration Molecular Mechanics Based on Combined Quantum Mechanical and Molecular Mechanical Calculations. *J. Chem. Theor. Comput.* **2006**, *2*, 1237–1254.
- (184) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Development of exchange-correlation functionals with minimal many-electron self-interaction error. *J. Chem. Phys.* **2007**, *126*, 191109.
- (185) Gaiduk, A. P.; Gustafson, J.; Gygi, F.; Galli, G. First-principles simulations of liquid water using a dielectric-dependent hybrid functional. *J. Phys. Chem. Lett.* **2018**, *9*, 3068–3073.
- (186) Willow, S. Y.; Zeng, X. C.; Xantheas, S. S.; Kim, K. S.; Hirata, S. Why is MP2-water “cooler” and “denser” than DFT-water? *J. Phys. Chem. Lett.* **2016**, *7*, 680–684.
- (187) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark calculations of noncovalent interactions of halogenated molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.
- (188) Rosenberg, R. E. Microsolvation of Fluoromethane. *J. Phys. Chem. A* **2016**, *120*, 7519–7528.
- (189) Rosenberg, R. E. The strength of hydrogen bonds between fluoro-organics and alcohols, a theoretical study. *J. Phys. Chem. A* **2018**, *122*, 4521–4529.
- (190) 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*, 16440–16453.
- (191) Evarestov, R. A. *Quantum Chemistry of Solids: The LCAO First Principles Treatment of Crystals*; Springer Science & Business Media, 2007; Vol. 153.
- (192) Hasnip, P. J.; Refson, K.; Probert, M. I.; Yates, J. R.; Clark, S. J.; Pickard, C. J. Density functional theory in the solid state. *Philos. Trans. R. Soc.* **2014**, *372*, 20130270.
- (193) 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.
- (194) Ayala, P.; Scuseria, G. Linear scaling second-order Møller–Plesset theory in the atomic orbital basis for large molecular systems. *J. Chem. Phys.* **1999**, *110*, 3660–3671.
- (195) Schutz, M.; Werner, H. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J. Chem. Phys.* **2001**, *114*, 661–681.
- (196) Flocke, N.; Bartlett, R. A natural linear scaling coupled-cluster method. *J. Chem. Phys.* **2004**, *121*, 10935–10944.
- (197) Distasio, R. A., Jr.; Steele, R. P.; Rhee, Y. M.; Shao, Y.; Head-Gordon, M. An improved algorithm for analytical gradient evaluation in resolution-of-the-identity second-order Møller–Plesset perturbation theory: Application to alanine tetrapeptide conformational analysis. *J. Comput. Chem.* **2007**, *28*, 839–856.
- (198) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]. *J. Chem. Phys.* **2018**, *148*, 011101.
- (199) Schutski, R.; Zhao, J.; Henderson, T. M.; Scuseria, G. E. Tensor-structured coupled cluster theory. *J. Chem. Phys.* **2017**, *147*, 184113.
- (200) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (201) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *J. Chem. Phys.* **1988**, *38*, 3098–3100.
- (202) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Nardelli, M. B.; 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.; Kucukbenli, E.; Lazzeri, M.; Marsili, M.; Marzari, N.; Mauri, F.; Nguyen, N. L.; Nguyen, H.-V.; Otero-de-la Roza, A.; Paulatto, L.; Ponce, 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.
- (203) Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918–1931.
- (204) Wolters, L. P.; Schyman, P.; Pavan, M. J.; Jorgensen, W. L.; Bickelhaupt, F. M.; Kozuch, S. The many faces of halogen bonding: a review of theoretical models and methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, S23–S40.
- (205) Jurečka, P.; Sponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
- (206) Manzhos, S.; Wang, X.; Dawes, R.; Carrington, T. A Nested Molecule-Independent Neural Network Approach for High-Quality Potential Fits. *J. Phys. Chem. A* **2006**, *110*, 5295–5304.
- (207) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.