

Reformulation of All ONIOM-Type Molecular Fragmentation Approaches and Many-Body Theories Using Graph-Theory-Based Projection Operators: Applications to Dynamics, Molecular Potential Surfaces, Machine Learning, and Quantum Computing

Published as part of *The Journal of Physical Chemistry A* virtual special issue “Krishnan Raghavachari Festschrift”.

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



Cite This: *J. Phys. Chem. A* 2024, 128, 466–478



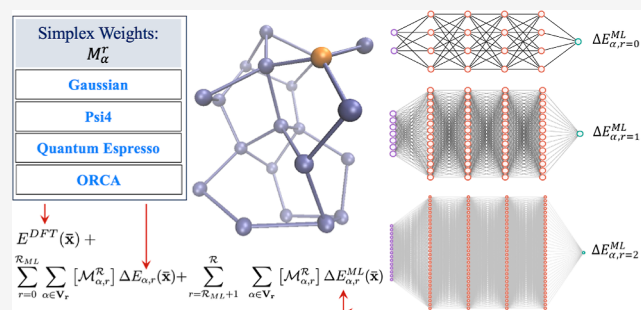
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: We present a graph-theory-based reformulation of all ONIOM-based molecular fragmentation methods. We discuss applications to (a) accurate post-Hartree–Fock AIMD that can be conducted at DFT cost for medium-sized systems, (b) hybrid DFT condensed-phase studies at the cost of pure density functionals, (c) reduced cost on-the-fly large basis gas-phase AIMD and condensed-phase studies, (d) post-Hartree–Fock-level potential surfaces at DFT cost to obtain quantum nuclear effects, and (e) novel transfer machine learning protocols derived from these measures. Additionally, in previous work, the unifying strategy discussed here has been used to construct new quantum computing algorithms. Thus, we conclude that this reformulation is robust and accurate.



1. INTRODUCTION

The accurate treatment of molecular properties often requires the correlated treatment of the electronic and nuclear degrees of freedom. However, even in the electronic structure, the size of systems that can be considered by standard post-Hartree–Fock approaches is strongly influenced by the intrinsic, steep, algebraic computational scaling of electron correlation methods as well as the number and quality of basis functions needed. For the study of electron correlation in most molecular systems, chemical accuracy may be achieved using the well-known CCSD(T) method¹ with an associated computational cost that scales as $O(N^7)$, where N represents the number of electronic basis functions. The quantum dynamical treatment of nuclei, on the other hand, is thought to be exponentially complex.^{2–5} As a result, the correlated study of molecular systems has several challenges. (a) The cost associated with obtaining accurate electronic potential surfaces needed to describe the nuclear degrees of freedom depends on the number of nuclear configurations needed to represent the potential surfaces; these may grow exponentially with the number of nuclear dimensions.^{6–11} (b) the storage and action of the operators such as the quantum propagator as well as nuclear wavepackets also may grow exponentially with

dimensions,^{12–15} (c) the intrinsic electronic structure for each nuclear configuration, as noted above, may itself be a steeply algebraic computational task for most chemical systems, and finally (d) as an intermediate, if classical description of nuclear degrees of freedom is sufficient, as would be the case of many chemical problems, even here the number of electronic structure calculations needed along with nuclear gradients becomes catastrophic as longer-time scale dynamics becomes a requirement to model many chemical and biological processes.

Despite algorithmic improvements for treating electron correlation,^{16–21} the post-Hartree–Fock electronic structure presents a significant challenge for systems of chemical interest. Density functional theory has allowed us to bridge this gap but significant challenges remain.^{22–24} The use of composite electronic energy expressions has grown significantly recently

Received: August 21, 2023

Revised: December 4, 2023

Accepted: December 8, 2023

Published: January 5, 2024



and involves the treatment of local interactions with post-Hartree–Fock methods.^{25–42} These are complementary to many-body methods^{43–51} that have also greatly contributed toward the development of reduced scaling methods.

In this paper, we focus on one family of molecular fragmentation methods that use ONIOM-type⁵² corrections to improve the computational scaling and accuracy of complex electronic structure calculations. Such fragmentation methods are closely connected to many-body treatment^{43,53,54} and local correlation methods^{27,51,55,56} and include the multicentered QM:QM formalism,^{38,57} the molecular tailoring approach (MTA),⁵⁸ ONIOM-XO,⁵⁵ the molecules-in-molecules (MIM) methodology,^{33,59–62} HMBI^{63,64} and the method of increments.^{65,66} The idea has been shown to be very powerful and perhaps the most impact from these methods has been due to Raghavachari's MIM approach that has been used to compute a variety of complex molecular properties including NMR coupling constants,^{67,68} vibrational circular dichroism,⁶⁹ and protein–ligand binding interactions,⁷⁰ at relatively low cost.

Here, we present an alternate reformulation of all ONIOM-type molecular fragmentation methods and many-body theories from a newly introduced graph-theoretic perspective,^{42,71–85} which allows us to derive new approaches for post-Hartree–Fock ab initio molecular dynamics^{42,71–74} for gas- and condensed-phase^{76,84} systems, computing high-quality molecular potential surfaces for quantum nuclear dynamics,^{75,78,79} efficient machine learning (ML) transfer learning protocols,^{81,84} new algorithms for quantum computing,^{80,83,85} and new algorithms for performing quantum nuclear dynamics with tensor networks.⁸² We briefly highlight some of these aspects in this publication.

The article is organized as follows: In Section 2, we present our graph-theory-based projection operator scheme, which divides a molecular Hamiltonian into several fragment Hamiltonians; critically, the projection provides an approximate resolution of identity using a set-theoretic or graph-theoretic procedure. Connections to the many-body theory are discussed in Section 2.1, and connections to other fragmentation procedures are described in Section 2.2. Studies describing extended Lagrangian- and Born–Oppenheimer-based molecular dynamics studies arising from the procedure given in Section 2 are described in Section 4, applications to molecular potential surfaces are provided in Section 3, and finally, ML applications are provided in Section 5. Conclusions are provided in Section 6.

2. GRAPH-THEORETIC DECOMPOSITION OF MOLECULAR SPACE DECOMPOSITION AND ASSOCIATED PROJECTION OPERATORS

We begin here with a set-theoretic decomposition of molecular space.⁸⁶ This approach is then adapted to a graph problem and used to decompose arbitrary molecular systems into a family of independent operations eventually leading to the many-body approximation.

We begin with a Venn diagram that divides a coordinate representation $\{|x\rangle\}$ into physical regions depicted here as A , B , and C . For our purposes, this coordinate representation $\{|x\rangle\}$ is essentially a basis representation of molecular space. We could have equally well replaced $\{|x\rangle\}$ with an atom-centered Gaussian basis set, which bears the signature of locality in molecular space, but for simplicity, we will retain $\{|x\rangle\}$. The regions, A , B , and C , may have overlaps and in Figure 1, we have superimposed the Venn diagram on a protonated water

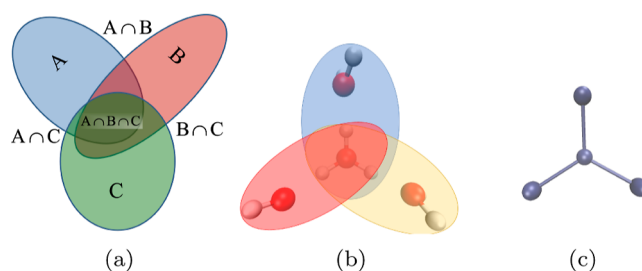


Figure 1. An illustration of the sets A , B , and C in (a) is superimposed on a protonated water cluster (b) that is then used to construct eq 1. A graph representation corresponding to (b) is shown in (c).

cluster. The Venn diagram divides the system into regions that may not, in general, be spatially orthogonal to each other, but as we will see below, this does not present a problem.

We then use the principle of inclusion–exclusion⁸⁶ from the set theory to resolve the identity that composes the Hilbert space, $\{|x\rangle\}$, as depicted within the Venn diagram

$$\begin{aligned} \mathbf{I} &\equiv \int_{A \cup B \cup C} dx |x\rangle \langle x| \\ &= \int_A dx |x\rangle \langle x| + \int_B dx |x\rangle \langle x| + \int_C dx |x\rangle \langle x| - \int_{A \cap B} dx |x\rangle \langle x| \\ &\quad - \int_{A \cap C} dx |x\rangle \langle x| - \int_{B \cap C} dx |x\rangle \langle x| + \int_{A \cap B \cap C} dx |x\rangle \langle x| \\ &= \mathcal{P}_A + \mathcal{P}_B + \mathcal{P}_C - \mathcal{P}_{A \cap B} - \mathcal{P}_{A \cap C} - \mathcal{P}_{B \cap C} + \mathcal{P}_{A \cap B \cap C} \end{aligned} \quad (1)$$

through integrals involving the dyadic terms, $|x\rangle \langle x|$, with domains within any chosen subset. Thus, we also introduce projection operators

$$\mathcal{P}_A \equiv \int_A dx |x\rangle \langle x| \quad (2)$$

for portions of the Hilbert space in eq 1. If atom-centered basis sets were used here instead of the $|x\rangle$ representation, this would complicate the discussion a little, given the non-orthogonality of atom-centered Gaussians. In that situation, one may need to resort to a Löwdin⁸⁷ or Cholesky⁸⁸ representation and the resultant basis functions may be considered to be some analogue of $|x\rangle$.⁸⁹

While eq 1 arises from the inclusion–exclusion in set theory,⁸⁶ an alternate approach is obtained by introducing a graph decomposition of the molecular structure. To begin with, a molecular assembly is partitioned into independent units that will be treated as nodes or vertices used to create a graph. These nodes may be determined based on chemical intuition or through numerical protocols. First-order interactions between these discrete units may be captured by creating edges that connect these nodes and may be considered as a union of atoms that are present within nodes.

Once the nodes and edges are defined, the system is represented as a graph, an illustration of which is provided in Figure 2.

The set of nodes described above is represented here as $\mathbf{V}_0$, and the family of edges is represented as $\mathbf{V}_1$. Consequently, these define a graph, $\mathcal{G} \equiv \{\mathbf{V}_0; \mathbf{V}_1\}$ (see Figure 2). However, now the graph also comprises higher order, rank- r objects, known as simplexes.^{90–92} Simplexes are geometric objects with an arbitrary number of vertices, where all pairs of vertices are

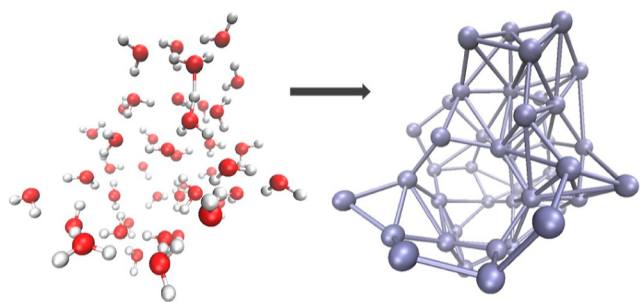


Figure 2. Protonated water is represented as a graph that acts as a distance-based truncation of many-body interactions.

connected.^{90–92} The set of such rank- r objects is represented as $\mathbf{V}_r$

$$\{\mathbf{V}_r | r = 0, 1, 2, \dots\} \equiv \{\mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2, \dots\} \quad (3)$$

and these capture higher-order interactions between the nodes.

An equivalent expression for the resolution of the identity in eq 1 may now be obtained in terms of projectors that encompass nodes, edges, and higher-order simplexes as

$$\begin{aligned} \mathbf{I} &= \sum_{\alpha \in \mathbf{V}_0} \mathcal{M}_\alpha^0 \mathcal{P}_{\alpha,0} - \sum_{\alpha \in \mathbf{V}_1} \mathcal{M}_\alpha^1 \mathcal{P}_{\alpha,1} + \sum_{\alpha \in \mathbf{V}_2} \mathcal{M}_\alpha^2 \mathcal{P}_{\alpha,2} - \dots \\ &= \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_r} \mathcal{M}_\alpha^r \mathcal{P}_{\alpha,r} \end{aligned} \quad (4)$$

and here

$$\mathcal{M}_\alpha^r \equiv \left[\sum_{m=r}^{\mathcal{R}} (-1)^m p_{\alpha}^{r,m} \right] \quad (5)$$

The quantity $p_{\alpha}^{r,m}$ is the number of times the α th rank- r term (in set $\mathbf{V}_r$) appears in all rank- m terms (in set $\mathbf{V}_m$), for $m \geq r$. Thus, $\mathcal{M}_\alpha^r$ is an overcounting correction for the number of times the α th rank- r term appears in objects of rank greater than or equal to r .

The parallels between eqs 4 and 1 may be explicated by rewriting eq 4 in decreasing order of rank, that is

$$\begin{aligned} \mathbf{I} &= (-1)^{\mathcal{R}} \left\{ \sum_{\alpha \in \mathbf{V}_{\mathcal{R}}} \mathcal{P}_{\alpha,\mathcal{R}} - \sum_{\alpha \in \mathbf{V}_{\mathcal{R}-1}} \mathcal{M}_\alpha^{\mathcal{R}-1} \mathcal{P}_{\alpha,\mathcal{R}-1} \right. \\ &\quad \left. + \sum_{\alpha \in \mathbf{V}_{\mathcal{R}-2}} \mathcal{M}_\alpha^{\mathcal{R}-2} \mathcal{P}_{\alpha,\mathcal{R}-2} - \sum_{\alpha \in \mathbf{V}_{\mathcal{R}-3}} \mathcal{M}_\alpha^{\mathcal{R}-3} \mathcal{P}_{\alpha,\mathcal{R}-3} + \dots \right\} \end{aligned} \quad (6)$$

where the appearance of alternating signs resembles that in eq 1. Additionally, for $\mathcal{R} = 1$, eq 6 becomes

$$\mathbf{I} = - \sum_{\alpha \in \mathbf{V}_1} \mathcal{P}_{\alpha,1} + \sum_{\alpha \in \mathbf{V}_0} \mathcal{M}_\alpha^0 \mathcal{P}_{\alpha,0} \quad (7)$$

which, for the graph in Figure 1c, leads to an identical result as in eq 1 but is easily applicable to more general situations such as in Figure 2.

2.1. Many-Body Theory from Eqs 1 and 4. We now begin with some molecular Hamiltonian $\mathcal{H}$ that is represented in the basis $\{|x\rangle\}$ and depicts the full molecular system, for example, those in the figures above. However, the complexity of such a system grows rapidly based on system size and this is not only depicted by the linear increase in the number of basis

elements within $\{|x\rangle\}$ but also by the correlations captured within $\mathcal{H}$, which increase in a steeply nonlinear and potentially exponential manner. To overcome this issue, we may apply the resolution of identity in eq 4 to $\mathcal{H}$ (or that defined using the set-theoretic expression in eq 1) to decompose $\mathcal{H}$ into a family of parallel systems given by

$$\begin{aligned} \mathbf{I}\mathcal{H} &= \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_r} \mathcal{M}_\alpha^r [\mathcal{P}_{\alpha,r} \mathcal{H}] \\ &= \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_r} \mathcal{M}_\alpha^r \mathcal{H}_{\alpha,r} \end{aligned} \quad (8)$$

where

$$\{\mathcal{H}_{\alpha,r} \equiv \mathcal{P}_{\alpha,r} \mathcal{H}\} \quad (9)$$

represent here a set of projected Hamiltonian, one for each molecular subsystem obtained from the simplex (α, r) within the graphical description. When a molecular system is divided using the graph, the set $\{\mathcal{H}_{\alpha,r}\}$ yields one Hamiltonian for each molecular fragment. This is illustrated in Figure 3.

We use the individual fragment molecular Hamiltonians, $\{\mathcal{H}_{\alpha,r}\}$, or suitable approximations to these, to obtain a family of fragment energies, for example, $\{E_{\alpha,r}^{\text{CCSD}}\}$, that, when used in eq 8, yields

$$E_R^{\text{graph-CCSD}}(\bar{x}) = \sum_{r=0}^{\mathcal{R}} \left\{ (-1)^r \sum_{\alpha \in \mathbf{V}_r} E_{\alpha,r}^{\text{CCSD}} \mathcal{M}_\alpha^r \right\}$$

Figure 3. Illustration of eqs 8 and 10. Using projection operators in eq 8, the Hamiltonian, $\mathcal{H}$, is represented as a family of independent fragment Hamiltonians, $\{\mathcal{H}_{\alpha,r}\}$, and these are processed in parallel as shown here.

$$E_{\mathcal{R}}^{\text{graph-CCSD}} = \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_r} E_{\alpha,r}^{\text{CCSD}} \mathcal{M}_{\alpha}^r \quad (10)$$

Equation 10 yields a stream of parallel computing processes that are completely independent of each other that each provides approximations to $\{E_{\alpha,r}^{\text{CCSD}}\}$. This idea is presented in Figure 3. Since these fragment Hamiltonians, $\{\mathcal{H}_{\alpha,r}\}$, are for much smaller as compared to the full system of interest, clearly, the computational burden reduces substantially. This has indeed been seen to be the case for several examples.

However, eq 10 is also closely related to the well-known many-body expansions,^{43,45–48} which can be clarified by writing out the appropriate form of eq 10 for $\mathcal{R} = 1$

$$E_{\mathcal{R}=1}^{\text{graph-CCSD}} = \sum_{\alpha \in \text{edges}} E_{\alpha,1} - \sum_{\alpha \in \text{nodes}} E_{\alpha,0} [p_{\alpha}^{0,1} - p_{\alpha}^{0,0}] \quad (11)$$

where $p_{\alpha}^{0,1}$ is the number of times the α th node (one-body term) appears in all edges (or two-body interactions) and $p_{\alpha}^{0,0}$ is the number of times node α appears in all nodes, that is, $p_{\alpha}^{0,0} = 1$. We may rewrite eq 11 as

$$E_{\mathcal{R}=1}^{\text{graph-CCSD}} = \sum_{\alpha \in \text{nodes}} E_{\alpha,0} + \left[\sum_{\alpha \in \text{edges}} E_{\alpha,1} - \sum_{\alpha \in \text{nodes}} p_{\alpha}^{0,1} E_{\alpha,0} \right] \quad (12)$$

These are essentially a one-body term that sums over all nodes

$$E_{1\text{-body}}^{\text{level,1}} \equiv \sum_{\alpha \in \text{nodes}} E_{\alpha,0} \quad (13)$$

and the two-body correction that is captured within the square-bracketed terms, $[\dots]$, in eq 12. However, eq 10 includes many-body contributions to arbitrary orders as can be seen by constructing a similar n -body analysis for three-body interactions (*faces*) where one may use $\mathcal{R} = 2$ to obtain

$$E_{\mathcal{R}=2}^{\text{graph-CCSD}} = \sum_{\alpha \in \text{nodes}} E_{\alpha,0} + \left[\sum_{\alpha \in \text{edges}} E_{\alpha,1} - \sum_{\alpha \in \text{nodes}} p_{\alpha}^{0,1} E_{\alpha,0} \right] + \left[\sum_{\alpha \in \text{faces}} E_{\alpha,2} - \sum_{\alpha \in \text{edges}} p_{\alpha}^{1,2} E_{\alpha,1} + \sum_{\alpha \in \text{nodes}} p_{\alpha}^{0,2} E_{\alpha,0} \right] \quad (14)$$

As for the case of eq 12, the last square-bracketed term in eq 14 captures the two-body corrections. Thus, eq 10 provides an adaptive MBE recipe to compute interactions to arbitrary order through an efficient graph-theoretic decomposition. A critical, and as yet unstated, underpinning of our graph-theoretic formalism is that simplexes are closed convex hulls.^{90,91} Thus, a higher rank simplex includes within it all components, lower-rank simplexes. For example, a rank-3 simplex has to be a tetrahedron and includes all four of its component triangles (rank-2 simplexes). These prescriptions directly follow from the fact that simplexes are closed convex hulls and the edges of any rank- r simplex are affinely independent.⁹² A rank- r simplex is constructed from the family linearly independent rank-1 simplexes (edges) $\{u_i\}$ as

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

where $\sum_{i=0}^r \lambda_i = 1$ and $\lambda_i \geq 0$.

This implies that the choice of spatial envelope which determines the maximum edge length within the graph is a way to control the maximum rank object considered. The requirement of simplexes being convex hulls is the single critical aspect that allows the mapping of our approach to the many-body theory. The absence of this situation will allow higher-order many-body interactions that do not then properly cancel the lower-order contributions and are not consistent with the set-theoretic inclusion–exclusion principle and many-body theory.^{42,86}

2.2. Improving Accuracy of MBE through ONIOM-Type Extrapolation. The description presented above provides a dynamic and flexible representation of local many-body interactions. We now discuss a composite energy measure^{25,93–96} that has been shown to converge faster as a function of maximum rank $\mathcal{R}$ ^{62,76,79} for ground state post-Hartree–Fock energies, AIMD trajectories, and multidimensional potentials. The energetic measure we begin with^{42,59,71–79,82} is a composite expression^{93–96} and consists of an ONIOM-type^{25,52} correction to a result from a lower level of theory

$$E_{\mathcal{R}}^{\text{extrap-CCSD}}(\bar{\mathbf{x}}) = E^{\text{DFT}}(\bar{\mathbf{x}}) + E_{\mathcal{R}}^{\text{graph-CCSD}} - E_{\mathcal{R}}^{\text{graph-DFT}} \quad (16)$$

where the term $E_{\mathcal{R}}^{\text{graph-CCSD}}$ is defined in eq 10 and $E_{\mathcal{R}}^{\text{graph-DFT}}$ is similarly defined as

$$E_{\mathcal{R}}^{\text{graph-DFT}} = \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_r} E_{\alpha,r}^{\text{DFT}} \mathcal{M}_{\alpha}^r \quad (17)$$

and hence

$$\begin{aligned} E_{\mathcal{R}}^{\text{extrap-CCSD}}(\bar{\mathbf{x}}) &= E^{\text{DFT}}(\bar{\mathbf{x}}) + \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_r} (E_{\alpha,r}^{\text{CCSD}} - E_{\alpha,r}^{\text{DFT}}) \mathcal{M}_{\alpha}^r \\ &= E^{\text{DFT}}(\bar{\mathbf{x}}) + \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_r} \Delta E_{\alpha,r} \mathcal{M}_{\alpha}^r \end{aligned} \quad (18)$$

where

$$\Delta E_{\alpha,r} = E_{\alpha,r}^{\text{CCSD}} - E_{\alpha,r}^{\text{DFT}} \quad (19)$$

Thus, the algorithm here envisions spawning out a family of computing processes, and this is shown in Figure 3. The resultant final energy in eq 18 is closely related to multiple ONIOM-based,^{29,38,52,55,56,59,97} molecular fragmentation methods^{27,29,32,40,41,47,48,51,98–104} of which MIM has proved to be remarkably versatile for a wide range of applications,^{33,60–62,105,106} as well as developments in the many-body theory.^{43,45–48,50,53,54,107–110} Equations 10, 16, and 17 have also been actively gas-phase and condensed-phase AIMD, multidimensional potential energy surfaces,^{42,71–79} and also provide new ways to construct training protocols in ML⁸¹ and for obtaining new quantum computing algorithms.^{80,85}

Finally, it is useful in illustrating the behavior of $\mathcal{M}'_\alpha$, which makes its appearance in eqs 10 and 18. As noted above, following eq 5, $\mathcal{M}'_\alpha$ is an overcounting correction for the number of times the α th rank- r term appears in all objects of rank greater than or equal to r . To illustrate this quantity, in Table 1, we present the behavior of $\mathcal{M}'_\alpha$ for various fragments

Table 1. $\mathcal{M}'_\alpha$ Values for H_3O^+ , H_5O_2^+ , H_7O_3^+ , and H_9O_4^+ from within a Protonated 21-Water Cluster System^a

$\mathcal{M}'_\alpha$ values for H_3O^+			
R	4.5 Å	7.0 Å	full graph
1	−5	−13	−19
2	5	67	171
3	−1	−184	−969
4	−1	301	3876

$\mathcal{M}'_\alpha$ values for H_5O_2^+			
R	4.5 Å	7.0 Å	full graph
1	1	1	1
2	−3	−9	−18
3	1	32	152
4	−1	−59	−817

$\mathcal{M}'_\alpha$ values for H_7O_3^+			
R	4.5 Å	7.0 Å	full graph
1	0	0	0
2	1	1	1
3	−2	−8	−17
4	−1	24	136

$\mathcal{M}'_\alpha$ values for H_9O_4^+			
R	4.5 Å	7.0 Å	full graph
1	0	0	0
2	0	0	0
3	1	1	1
4	1	−5	−16

^aThe respective graphical depictions are shown in Figure 4.

as we increase $\mathcal{R}$ in eqs 10 and 18 and also the maximum edge length allowed for the creation of the graph. The graphical representations that lead to these values are given in Figure 4. Clearly, as the value of $\mathcal{R}$ is increased, more and more many-body interactions, beyond the closest bonding interactions, and also longer-range weaker interactions are included within the approximation and the value of $\mathcal{M}'_\alpha$ grows significantly.

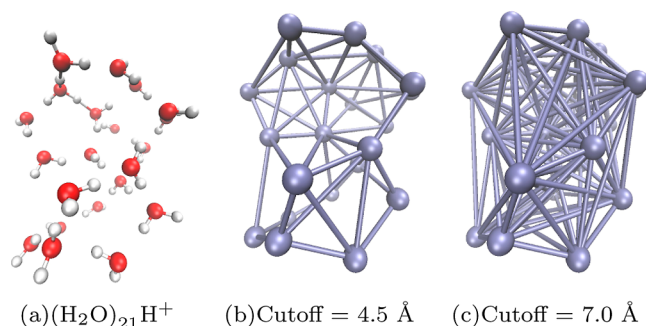


Figure 4. Graphical representation of a protonated water cluster with varied edge length cutoffs. The choice of edge-cutoff impacts the weights, $\mathcal{M}'_\alpha$, of the individual components used in eq 18. The weights of example simplexes are shown in Table 1.

3. MOLECULAR POTENTIAL SURFACES FROM EQ 16

In refs 42, 71–76, we have discussed the efficient evaluation of correlation energy and basis-set extrapolation using the formalism above, but the graphs that are defined based on instantaneous molecular geometries connectivities may change when atoms move during dynamics or potential energy surface calculations. This may create singularities in potential surfaces and in refs 75, 78, 82, a new approach is introduced to compute potential surfaces using a weighted set of graphs. For one geometry, multiple representations of eq 16 may be considered depending on the graphical representation of the system such that $\bar{\mathbf{x}}$ and $\mathcal{G}_\beta$ together map to provide an estimate for eq 16, $E_{\mathcal{R},\beta}^{\text{extrap-CCSD}}(\bar{\mathbf{x}})$, that is

$$(\bar{\mathbf{x}}, \mathcal{G}_\beta) \rightarrow E_{\mathcal{R},\beta}^{\text{extrap-CCSD}}(\bar{\mathbf{x}}; \mathcal{G}_\beta) \quad (20)$$

When considering a family of such graphical representations, $\{\mathcal{G}_\beta\}$, with respective maximum ranks, $\mathcal{R}_\beta$, the energy of the system becomes a probabilistic sum over $\{E_{\mathcal{R},\beta}^{\text{extrap-CCSD}}(\bar{\mathbf{x}}; \mathcal{G}_\beta)\}$. The graphs considered can be thought of as “valence bond”^{111–113} constructs or “diabatic states,”^{114–121} reminiscent of nonadiabatic electronic structure theory.¹²² The energy of the system from multiple graphs (or a single hyper graph^{91,123}) is obtained from a probabilistic sum

$$\langle E(\bar{\mathbf{x}}) \rangle = \sum_{\beta} \rho_{\beta}(\bar{\mathbf{x}}) \cdot E_{\mathcal{R},\beta}^{\text{extrap-CCSD}}(\bar{\mathbf{x}}; \mathcal{G}_\beta) \quad (21)$$

In refs 75, 78, and 82, we introduce a variational procedure to obtain $\rho_{\beta}(\mathbf{R})$. Generalizations to multiple dimensions are discussed in refs 78 and 82.

3.1. Computing Smooth Potentials in AIMD in a Fashion Inspired by Eq 21. The use of dynamic and flexible representations of local many-body interactions for the computing of potential surfaces for AIMD is known to yield discontinuities in energy and forces. This problem is quite similar to that of the full potential energy surfaces but with the caveat that the smoothing on the surfaces must be done on the fly in a more adaptive manner following the dynamic evolution of the molecular degrees of freedom during the trajectory. To achieve such a dynamic smoothening of the potential energy surface, one may obtain, just one would in eq 21, a family of graphs $\{\mathcal{G}_\beta\}$. Each graph provides a different set of energy and gradients, then this family of graphs are combined to obtain the overall energy by

$$E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\bar{\mathbf{x}}) = E^{\text{DFT}}(\bar{\mathbf{x}}) + \sum_{\beta} v_{\beta}(\bar{\mathbf{x}}) \left(\sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_r^{\beta}} \Delta E_{\alpha,r,\beta}(\bar{\mathbf{x}}) \mathcal{M}'_{\alpha,\beta} \right) \quad (22)$$

where the square-bracketed term (...) is the substitution of eq 19 into eq 17. This combination of terms is added to the full-system DFT leading to an equation analogous to eq 18, with the distinction that the terms $\Delta E_{\alpha,r,\beta}(\bar{\mathbf{x}})$ and $\mathcal{M}'_{\alpha,\beta}$ now have $\{\mathcal{G}_\beta\}$ dependence. This approach differs from eq 21 where the energy of the system is a probabilistic sum over multiple fragmentation topologies (or graphs); eq 22, by contrast, uses a numerical weighting scheme to obtain smooth energies and gradients as discussed in ref 79.

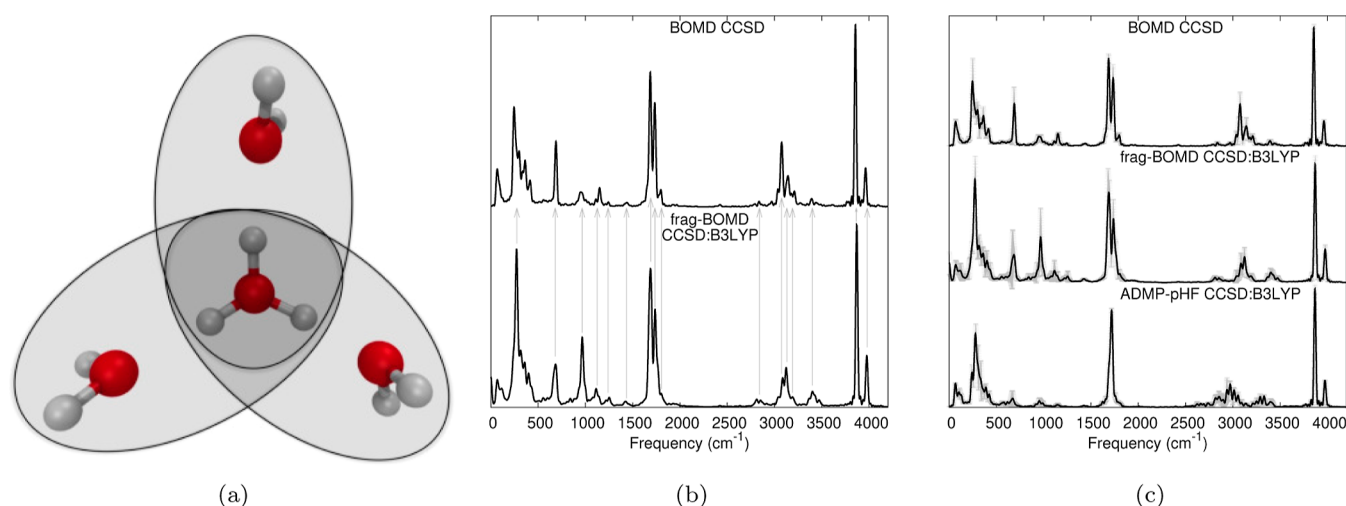


Figure 5. Vibrational density of states (b,c) for the Eigen (a) protonated water clusters are presented at the CCSD level of accuracy.

4. BORN–OPPENHEIMER AND EXTENDED LAGRANGIAN APPROACHES USING EQ 16

The steep computational scaling needed for accurate treatment of molecular properties is especially an issue for AIMD and determination of quantum nuclear effects. In a series of publications,^{42,71–76} we have shown how eq 16 is to approximate potential surfaces and AIMD trajectories. Both extended Lagrangian- and Born–Oppenheimer-based ab initio molecular dynamics simulations are performed with CCSD and MP2 accuracy but at DFT cost.^{42,71–73} Hence, for the first time, in refs 71 and 72, we provided Car–Parrinello dynamics, with CCSD accuracy. In ref 74, we have also demonstrated that weak interactions (such as in hydrogen bonds) can be accurately studied and approximations to large-basis AIMD trajectories, such as 6-311++G(2df,2pd), can be constructed through computational effort commensurate with smaller basis sets, such as 6-31+G(d).

Toward AIMD, the nuclear gradients for the energy in eq 16 may be written as

$$\frac{\partial E_{\mathcal{R}}^{\text{extrap-CCSD}}(\bar{\mathbf{x}})}{\partial \bar{\mathbf{x}}} = \frac{\partial E^{\text{DFT}}(\bar{\mathbf{x}})}{\partial \bar{\mathbf{x}}} + \sum_{r=0}^{\mathcal{R}} (-1)^r \left\{ \sum_{\alpha \in \mathbf{V}_r} \left(\frac{\partial E_{\alpha,r}^{\text{CCSD}}}{\partial \bar{\mathbf{x}}_{\alpha,r}} - \frac{\partial E_{\alpha,r}^{\text{DFT}}}{\partial \bar{\mathbf{x}}_{\alpha,r}} \right) \left(\frac{\partial \bar{\mathbf{x}}_{\alpha,r}}{\partial \bar{\mathbf{x}}} \right) \mathcal{M}_{\alpha}^r \right\} \quad (23)$$

The nuclear coordinates for the molecular fragment representing the α th r -rank many-body term, $\bar{\mathbf{x}}_{\alpha,r}$, may not be entirely a subset of the system coordinates, $\bar{\mathbf{x}}$, as it may include link atoms if bonds are broken in the formation of the nodal definitions in the graph, as allowed by ONIOM.⁵² If link atoms are included then $\left[\frac{\partial \bar{\mathbf{x}}_{\alpha,r}}{\partial \bar{\mathbf{x}}} \right]$ is a Jacobian needed to transform the r -rank many-body gradients back to the full system gradients.^{42,124} These gradients enable Born–Oppenheimer molecular dynamics using eq 16, as demonstrated in refs 42, 71–74.

As the size of the system increases, the full-system gradients $\left[\frac{\partial E^{\text{DFT}}(\bar{\mathbf{x}})}{\partial \bar{\mathbf{x}}} \right]$ dominate the AIMD computational costs. In refs 71, 72, and 74, an r -rank many-body extended Lagrangian^{125,126} is introduced where the electronic single-particle density matrix, $\mathbf{P}$, that depicts the energy $E^{\text{DFT}}(\bar{\mathbf{x}})$ in eq 16, are treated as dynamical variables. The electronic parameters, $\mathbf{P}$, that lead to $E^{\text{DFT}}(\bar{\mathbf{x}})$ are propagated with the nuclear variables through an adjustment of the relative time scales between the full-system, low-level treatment and nuclear degrees of freedom. This is essentially a Car–Parrinello-like method¹²⁷ but is implemented using the atom-centered Gaussian basis functions and single particle density matrices that determine and hence follow the atom-centered density matrix propagation (ADMP)^{128–131} protocol. This treatment allows the ability to calculate post-Hartree–Fock-based extended Lagrangian (Car–Parrinello-like) trajectories. This methodology is termed “Atom-centered Density Matrix Propagation with post-Hartree–Fock accuracy” and abbreviated, ADMP-pHF.^{71,72} The associated extended Lagrangian that serves for post-Hartree–Fock accuracy as well as basis-set extrapolated dynamics is given by

$$\mathcal{L} = \frac{1}{2} \text{Tr}[\mathbf{V}^T \mathbf{M} \mathbf{V}] + \frac{1}{2} \text{Tr}[(\mu^{1/4} \mathbf{W} \mu^{1/4})^2] - E_{\mathcal{R}}^{\text{extrap-CCSD}}(\bar{\mathbf{x}}) - \text{Tr}[\Lambda(\mathbf{P}^2 - \mathbf{P})] \quad (24)$$

Here the parameters $\bar{\mathbf{x}}$ and $\mathbf{V}$ represent the classical nuclear positions and velocities, with masses, $\mathbf{M}$. The velocity Verlet¹³² integration method is used to compute the dynamics of $\{\bar{\mathbf{x}}, \mathbf{V}; \mathbf{P}, \mathbf{W}\}$. In refs 42, 71–74, the effectiveness of eq 16 for the construction of dynamics trajectories at post-Hartree–Fock accuracy is demonstrated by studying protonated water clusters and polypeptide fragments. The efficiency and accuracy of these trajectories are evaluated in refs 42, 71–74 in multiple ways. First, given the Hamiltonian nature of Born–Oppenheimer and ADMP-pHF, total energy conservation and the drift in total energy are both evaluated in detail in refs 42, 71–74. The total energy is conserved to within fractions of kcal/mol for all studies, and the energy drift is also found to be similarly small. The accuracy of the vibrational density of states obtained through these trajectories is gauged by comparison with classical trajectories at the higher level, namely, CCSD or MP2. (see Figure 5.) The vibrational density of states is a

spectral (Fourier) representation of the trajectory, which partitions the nuclear velocity contributions as a function of frequency. The comparison of these spectral features is used to gauge the effectiveness of the low-cost, graph-theory-based many-body expansions used within AIMD as compared to standard AIMD treatments. In Figure 5, we provide one specific illustration of such a comparison with several others in refs 42, 71–74. The arrows in Figure 5b demonstrate the quality of the fragmentation-based AIMD and ADMP-pHF trajectories in computing accurate vibrational modes.

5. ML METHODS TO COMPUTE $\Delta E_{\alpha,r}$ IN EQ 18

ML approaches¹³³ have recently become popular in quantum chemistry, with a particular increase in interest following Google DeepMind's,^{134,135} efforts toward protein structure prediction. Indeed, DeepMind aims to construct an extremely complex neural network arising from experimental data to optimize and predict the molecular structure in very large systems. As a result of such an exciting development, a variety of ML techniques such as neural networks (NN)¹³⁶ and Gaussian process regression¹³⁷ have been used for computing accurate potential energy surfaces.^{138–144} The key idea here is that as long as a reasonable training set can be assembled, a potentially nonlinear extrapolation scheme can be constructed, presumably using NN, that can provide reasonable accuracy. One of the key issues that affect the training process in potential energy surface calculations is the fact that the size of a suitable training data set may increase drastically with system size.^{136,145–147} In refs 81 and 84, we create a family of neural networks that each yield an estimate for eq 18 given by

$$\Delta E_{\alpha,r} \xrightarrow{\text{ML}} \Delta E_{\alpha,r}^{\text{ML}} \quad (25)$$

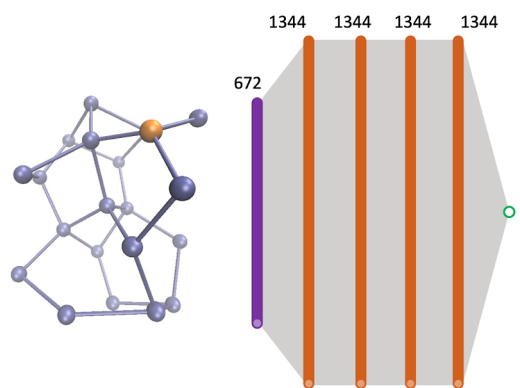
and thus

$$\begin{aligned} E_{\mathcal{R},\text{ML}}^{\text{extrap-CCSD}(\bar{\mathbf{x}})} &= E^{\text{DFT}}(\bar{\mathbf{x}}) + \sum_{r=0}^{\mathcal{R}} (-1)^r \\ &\quad \sum_{\alpha \in \mathcal{V}_r} (E_{\alpha,r}^{\text{CCSD}} - E_{\alpha,r}^{\text{DFT}}) \mathcal{M}_{\alpha}^r \\ &= E^{\text{DFT}}(\bar{\mathbf{x}}) + \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathcal{V}_r} \Delta E_{\alpha,r}^{\text{ML}} \mathcal{M}_{\alpha}^r \end{aligned} \quad (26)$$

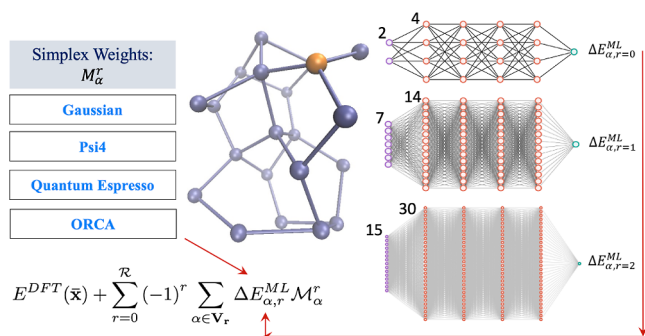
That is, ML estimates, $\Delta E_{\alpha,r}^{\text{ML}}$, replace $\Delta E_{\alpha,r}$ in eq 16. As shown in Figure 7 and in refs 81 and 84, training models can be constructed to reproduce AIMD data with roughly 10% of the effort needed as compared to that for the full fragmentation calculations. The idea is depicted in Figure 6. As can be seen in Figure 7, the density of the NN grows significantly as system size grows, and hence, in essence, eq 26 allows one to decouple parts of the NN for the full system, based on graph-theory-based molecular fragmentation to arrive at the needed family of networks. Furthermore, the individual networks being far smaller in size are easier to train and additionally, if chosen in a careful way can have a more universal application, beyond the full system used to obtain the training set.

To quantify exactly the change in complexity, we first note that these NNs essentially perform the following operations sequentially

$$\mathbf{x}_{a+1} = f_{a+1}(W_{a,a+1} \cdot \mathbf{x}_a) \quad (27)$$



(a) Full system neural network: Number of circles are listed and determine the complexity of model generation



(b) Fragment neural networks: Far fewer neurons and hence much reduced complexity in determining an accurate model

Figure 6. Comparison between (a) direct full-system learning and (b) ML process following graph-theoretic fragmentation. Every purple circle in the input layer of NNs represents 3 features, and every orange circle in the hidden layer represents 6 neurons. These numbers are chosen as a cofactor across all fragments to simplify the figure. Purely based on visual inspection and the number of neurons depicted, it is clear that the complexity of model generation in (a) is far more complex and almost intractable as compared to that in (b).

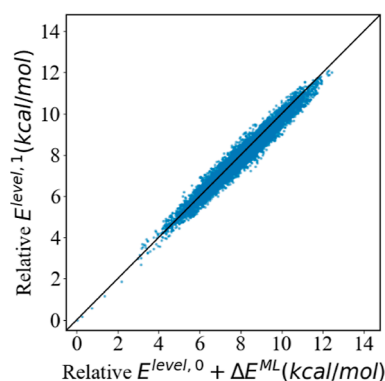


Figure 7. Correlation between ML approximations and graph-theoretic molecular fragmentation energy, that is, eq 26 and the full-system high-level energy for a range of solvated Zundel geometries obtained from AIMD calculations.

where f_{a+1} represents the activation function for the $(a + 1)$ -th layer, and $\{W_{a,a+1}\}$ are the weight matrices (including bias) connecting the a -th and $(a + 1)$ -th layers. The operation $W_{a,a+1} \cdot \mathbf{x}_a$ is essentially a matrix-vector operation. As part of the training process, it is necessary to find the terms in $\{W_{a,a+1}\}$, for all a , which then defines the model. Thus, in some sense,

the training process helps compute the unknown terms in $\{W_{a,a+1}\}$ and the complexity of any NN may be quantified in terms of the number of unknown weights that need to be obtained, that is

$$N_i \times N_h + (M_h - 1) \times (N_h)^2 + N_h \quad (28)$$

where N_i is the number of nodes in the input layer, and N_h is the number of elements in the distance matrix for a given molecule. Thus, $N_i \equiv N_{\text{Atoms}} \times (N_{\text{Atoms}} + 1)/2 - N_{\text{Atoms}}$. Here, the quantity N_{Atoms} is the number of atoms in the system. The quantity N_h is the number of nodes in each hidden layer, and M_h is the number of hidden layers. This expression, of course, assumes that the network is completely connected between adjacent layers. This complexity will reduce if, for example, a convolution network¹⁴⁸ is chosen, but the advantages noted below apply in general to the formalism noted here. In ref 81, we choose $N_h = 4N_i$ and hence the complexity of the network and the number of parameters in it grow as

$$\begin{aligned} \eta &= 4 \times N_i^2 + 16 \times (M_h - 1) \times (N_i)^2 + 4 \times N_i \\ &= N_{\text{Atoms}}^2 (N_{\text{Atoms}} - 1)^2 + 4(M_h - 1)N_{\text{Atoms}}^2 \\ &\quad (N_{\text{Atoms}} - 1)^2 + 2N_{\text{Atoms}}(N_{\text{Atoms}} - 1) \\ &\rightarrow O(N_{\text{Atoms}}^4) \end{aligned} \quad (29)$$

Thus, as seen through the illustration in Figure 6a for the protonated 21-water cluster, when the full system is used to compute a ML model, the number of terms in the input layer is already large (64 atoms and hence $N_i = 64 \times 63/2 = 2016$ and is represented by $2016/3 = 672$ circles in the figure where 3 is a common factor to simplify plotting for all fragments and the full system). Table 2 complements the discussion here. Correspondingly, the number of terms in the hidden layer is $N_h = 4 \times 2016 = 8064$. Consequently, the number of weights in the single NN, with four hidden layers as in Figure 6a, is $2016 \times 8064 + 3 \times 8064^2 + 8064 = 2 \times 10^8$ terms. See Table 2. The resultant optimization problem is hard because one needs to find a single solution to this 200 million-dimensional nonlinear problem.

By contrast, eq 26 produces a family of independent NNs, as seen in Figure 6b, and the number of terms in these NNs can be seen to be significantly smaller purely by visual inspection of Figure 6b. When 3-body terms ($R = 2$) are included, the number of weights in the NN corresponding to each graph node (top row network of Figure 6b), with 3–4 atoms per node (3 atoms for water and 4 for H_3O^+ , and hence $N_i = 3$ for water and $N_i = 6$ for H_3O^+), is roughly $4 \times 6^2 + 16 \times 3 \times 6^2 + 4 \times 6$ which is 1896 for H_3O^+ . Table 2 provides a more complete discussion. For each independent edge (6–7 atoms), the number of weights in one NN is $4 \times 21^2 + 16 \times 3 \times 21^2 + 4 \times 21$ which is of the order of 23,000 for H_5O_2^+ . Thus, between nodes and edges, there is a substantial reduction in the complexity of the NNs as compared to the full system, making the training process easier. For three-body interactions that are included in each independent face NN, the number of parameters is $4 \times 45^2 + 16 \times 3 \times 45^2 + 4 \times 45$ which is around 10^5 for H_7O_3^+ . The highest complexity here clearly arises from the face NNs, but this is nearly 3 orders of magnitude lower in complexity with respect to the NN for the full system. See Table 2, but this can also be seen visually from the density of NN nodes in Figure 6a as compared to that in Figure 6b.

Table 2. Reduced Complexity of NNs^a

	Figure 6a		Figure 6b	
	$(\text{H}_2\text{O})_{21}\text{H}^+$		$(\text{H}_2\text{O})_n\text{H}^+{}^b$	$(\text{H}_2\text{O})_n{}^c$
$N_i{}^e$	2016	nodes ^d	6	3
		edges ^f	21	15
		faces ^g	45	36
$N_h{}^h$	8064	nodes ^d	24	12
		edges ^f	84	60
		faces ^g	180	144
$M_h{}^i$	4		4	4
η^j	2×10^8		$\leq 1.3 \times 10^{5k}$	$\leq 8 \times 10^{4k}$

^aThe key idea is that, as the system size grows, the complexity of constructing a NN model grows prohibitively and beyond that mentioned under the column labeled “Figure 6a”. See eq 27. However, for systems where 3- and 4-body interactions are sufficient (most chemical systems), the scaling will remain tractable based on the description under the column labelled “Figure 6b”. The table here complements Figure 6, where the number of nodes in the figure has been uniformly scaled down for clarity. ^bProtonated fragments. ^cNeutral fragments. ^dNodes: single water or hydronium. ^eNumber of input nodes in eq 28. These numbers are scaled by 1/3 for pictorial clarity in Figure 6. ^fEdges: water-dimer or Zundel cation. ^gFaces: water-trimer or Eigen cation. ^hNumber of nodes in each hidden layer in eq 28. These numbers are scaled by 1/6 in Figure 6 to maintain pictorial clarity. ⁱNumber of hidden layers in eq 28. ^jNetwork complexity given by the total number of unknown variables needed to create NN models. See eq 28 and the associated discussion. ^kThe number of parameters here is due to faces. The number is less for nodes and edges. See text.

However, when the system size grows, the complexity of a ML protocol constructed for the full system, represented by the second row in Table 2, continues to grow according to eq 27. By contrast, when eq 26 is used, the complexity grows as

$$\begin{aligned} \eta_R &= 4 \times N_{i,R}^2 + 16 \times (M_h - 1) \times (N_{i,R})^2 + 4 \times N_{i,R} \\ &= N_{\text{Atoms}}^R{}^2 (N_{\text{Atoms}}^R - 1)^2 + 4(M_h - 1) \\ &\quad N_{\text{Atoms}}^R{}^2 (N_{\text{Atoms}}^R - 1)^2 + 2N_{\text{Atoms}}^R(N_{\text{Atoms}}^R - 1) \\ &\rightarrow O(N_{\text{Atoms}}^R{}^4) \end{aligned} \quad (30)$$

where N_{Atoms}^R is the number of atoms in the maximum rank R simplex. It is expected that for large systems $N_{\text{Atoms}}^R \ll N_{\text{Atoms}}$, the total number of atoms in the full system. The algorithm has been demonstrated for surface–adsorbate interactions in ref 84, for protonated water clusters in ref 81, and in Figure 7, where the horizontal axis refers to results from eq 26, whereas the vertical axes refer to the energy value at CCSD. Clearly, the agreement is very high between the fragment-based ΔML model and the more expensive higher-level electronic structure treatment.

6. CONCLUSIONS

We have discussed a graph-theory-based molecular fragmentation procedure and embedding procedure with strong connections to many-body theory, the molecules-in-molecules procedure, and many other fragmentation approaches. The many-body approximation is adaptively computed using rank- r graph-theoretic simplexes composed from a power set of “coarse-grained” local partitions (nodes) in a chemical system; the higher rank simplexes capture increasingly nonlocal

interactions between the “coarse-grained” units. The graph-theoretic many-body expression introduced here perturbatively improves on a full-system calculation conducted at an affordable lower level of theory and basis as in ONIOM.

This method allows access to higher-quality (post-Hartree–Fock) electronic structure methodologies at a lower computational cost. The approach has been used to (a) efficiently compute post-Hartree–Fock Born–Oppenheimer^{42,72–74} and extended Lagrangian^{71–74} AIMD trajectories, (b) obtain multidimensional potential energy surfaces for the treatment of nuclear quantum effects where the surfaces are obtained at post-Hartree–Fock accuracy,^{75,78} (c) provide efficient ML protocols,^{81,84} (d) derive efficient methods for tensor network-based quantum nuclear dynamics strategies,⁸² and (e) obtain efficient, reduced quantum circuit depth algorithms for quantum computing.^{80,83,85} The cost reduction, robustness, and accuracy demonstrations render great promise to the methods discussed here.

AUTHOR INFORMATION

Corresponding Author

Srinivasan S. Iyengar – Department of Chemistry, Department of Physics, and the Indiana University Quantum Science and Engineering Center (IU-QSEC), Indiana University, Bloomington, Indiana 47405, United States; orcid.org/0000-0001-6526-2907; Email: iyengar@indiana.edu

Authors

Timothy C. Ricard – Department of Chemistry, Department of Physics, and the Indiana University Quantum Science and Engineering Center (IU-QSEC), Indiana University, Bloomington, Indiana 47405, United States

Xiao Zhu – Department of Chemistry, Department of Physics, and the Indiana University Quantum Science and Engineering Center (IU-QSEC), Indiana University, Bloomington, Indiana 47405, United States

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpca.3c05630>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation grants CHE-2102610 to author S.S.I.

REFERENCES

- (1) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- (2) Feynman, R. P.; Hibbs, A. R. *Quantum Mechanics and Path Integrals*; McGraw-Hill Book Company: New York, 1965.
- (3) Meyer, H.-D.; Manthe, U.; Cederbaum, L. S. The multi-configurational time-dependent Hartree approach. *Chem. Phys. Lett.* **1990**, *165*, 73–78.
- (4) Nielsen, M. A.; Chuang, I. L. *Quantum Computation and Quantum Information*; Cambridge University Press: Cambridge, 2000.
- (5) Feynman, R. P.; Hey, J.; Allen, R. W. *Feynman Lectures on Computation*; Addison-Wesley Longman Publishing Co., Inc., 1998.
- (6) Jakowski, J.; Sumner, I.; Iyengar, S. S. Computational improvements to quantum wave packet *ab initio* molecular dynamics using a potential-adapted, time-dependent deterministic sampling technique. *J. Chem. Theory Comput.* **2006**, *2*, 1203–1219.
- (7) Sumner, I.; Iyengar, S. S. Quantum Wavepacket *Ab Initio* Molecular Dynamics: An Approach for Computing Dynamically Averaged Vibrational Spectra Including Critical Nuclear Quantum Effects. *J. Phys. Chem. A* **2007**, *111*, 10313–10324.
- (8) Braams, B. J.; Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577–606.
- (9) Hocker, D.; Li, X.; Iyengar, S. S. Shannon entropy based time-dependent deterministic sampling for efficient “on-the-fly” quantum dynamics and electronic structure. *J. Chem. Theory Comput.* **2011**, *7*, 256–268.
- (10) Peláez, D.; Meyer, H.-D. The multigrid POTFIT (MGPF) method: Grid representations of potentials for quantum dynamics of large systems. *J. Chem. Phys.* **2013**, *138*, 014108.
- (11) DeGregorio, N.; Iyengar, S. S. Efficient and Adaptive Methods for Computing Accurate Potential Surfaces for Quantum Nuclear Effects: Applications to Hydrogen- Transfer Reactions. *J. Chem. Theory Comput.* **2018**, *14*, 30–47.
- (12) DeGregorio, N.; Iyengar, S. S. Adaptive Dimensional Decoupling for Compression of Quantum Nuclear Wave Functions and Efficient Potential Energy Surface Representations through Tensor Network Decomposition. *J. Chem. Theory Comput.* **2019**, *15*, 2780–2796.
- (13) Hackbusch, W.; Khoromskij, B. N. Tensor-product approximation to operators and functions in high dimensions. *J. Complex.* **2007**, *23*, 697–714.
- (14) Greene, S. M.; Batista, V. S. Tensor-Train Split-Operator Fourier Transform (TT-SOFT) Method: Multidimensional Non-adiabatic Quantum Dynamics. *J. Chem. Theory Comput.* **2017**, *13*, 4034–4042.
- (15) Orús, R. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Ann. Phys.* **2014**, *349*, 117–158.
- (16) 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.
- (17) 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.
- (18) Flocke, N.; Bartlett, R. A natural linear scaling coupled-cluster method. *J. Chem. Phys.* **2004**, *121*, 10935–10944.
- (19) Distasio, R. A.; 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.
- (20) 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.
- (21) Schutski, R.; Zhao, J.; Henderson, T. M.; Scuseria, G. E. Tensor-structured coupled cluster theory. *J. Chem. Phys.* **2017**, *147*, 184113.
- (22) 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.
- (23) 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.
- (24) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289–320.
- (25) 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.
- (26) 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.
- (27) Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.* **2015**, *115*, 5607–5642.
- (28) 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.
- (29) Beran, G. J. O. Modeling polymorphic molecular crystals with electronic structure theory. *Chem. Rev.* **2016**, *116*, 5567–5613.
- (30) 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.
- (31) 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.
- (32) 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.
- (33) Thapa, B.; Beckett, D.; Jovan Jose, K. V.; Raghavachari, K. Assessment of Fragmentation Strategies for Large Proteins Using the Multilayer Molecules-in-Molecules Approach. *J. Chem. Theory Comput.* **2018**, *14*, 1383–1394.
- (34) König, C.; Christiansen, O. Linear-scaling generation of potential energy surfaces using a double incremental expansion. *J. Chem. Phys.* **2016**, *145*, 064105.
- (35) 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.
- (36) Mande, S.; Jose, K. V. J. A new model system preparation approach for ONIOM: Accurate estimation of energies and forces of molecules with pi-bonds. *J. Indian Chem. Soc.* **2019**, *96*, 1003.
- (37) Fang, T.; Li, Y.; Li, S. Generalized energy-based fragmentation approach for modeling condensed phase systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *7*, No. e1297.
- (38) 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.
- (39) Fedorov, D. G.; Kitaura, K. Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.
- (40) 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.
- (41) 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.
- (42) 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.
- (43) 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.
- (44) Nagayoshi, K.; Ikeda, T.; Kitaura, K.; Nagase, S. *J. Theory Comput. Chem.* **2003**, *02*, 233–244.
- (45) 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.
- (46) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many Body Expansion for Simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1–6.
- (47) 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.
- (48) 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.
- (49) 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.
- (50) 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.
- (51) Herbert, J. M. Fantasy versus reality in fragment-based quantum chemistry. *J. Chem. Phys.* **2019**, *151*, 170901.
- (52) Maseras, F.; Morokuma, K. IMOMM: A new integrated *ab initio* + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* **1995**, *16*, 1170–1179.
- (53) Murrell, J.; Carter, S.; Farantos, S.; Huxley, P.; Varandas, A. *Molecular Potential Energy Functions*; Wiley: New York, 1984.
- (54) Varandas, A.; Pais, A. A realistic double many-body expansion (DMBE) potential energy surface for ground-state O_3 from a multiproperty fit to *ab initio* calculations, and to experimental spectroscopic, inelastic scattering, and kinetic isotope thermal rate data. *Mol. Phys.* **1988**, *65*, 843–860.
- (55) 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.
- (56) 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.
- (57) Hopkins, B. W.; Tschumper, G. S. Integrated quantum mechanical approaches for extended π systems: Multicentered QM/QM studies of the cyanogen and diacetylene trimers. *Chem. Phys. Lett.* **2005**, *407*, 362–367.
- (58) Sahu, N.; Yeole, S. D.; Gadre, S. R. Appraisal of molecular tailoring approach for large clusters. *J. Chem. Phys.* **2013**, *138*, 104101.
- (59) 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.
- (60) Sadhukhan, T.; Beckett, D.; Thapa, B.; Raghavachari, K. Coupling Constants, High Spin, and Broken Symmetry States of Organic Radicals: an Assessment of the Molecules-in-Molecules Fragmentation-Based Method. *J. Chem. Theory Comput.* **2019**, *15*, 5998–6009.
- (61) Thapa, B.; Beckett, D.; Erickson, J.; Raghavachari, K. Theoretical Study of Protein-Ligand Interactions Using the Molecules-in-Molecules Fragmentation-Based Method. *J. Chem. Theory Comput.* **2018**, *14*, 5143–5155.
- (62) Debnath, S.; Sengupta, A.; Jose, K. V. J.; Raghavachari, K. Fragment-Based Approaches for Supramolecular Interaction Energies: Applications to Foldamers and Their Complexes with Anions. *J. Chem. Theory Comput.* **2018**, *14*, 6226–6239.
- (63) Sontising, W.; Beran, G. J. O. Theoretical assessment of the structure and stability of the λ phase of nitrogen. *Phys. Rev. Mater.* **2019**, *3*, 095002.
- (64) 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.
- (65) Stoll, H.; Paulus, B.; Fulde, P. On the accuracy of correlation-energy expansions in terms of local increments. *J. Chem. Phys.* **2005**, *123*, 144108.
- (66) Paulus, B. The method of increments - a wavefunction-based *ab initio* correlation method for solids. *Phys. Rep.* **2006**, *428*, 1–52.
- (67) Jose, K. V. J.; Raghavachari, K. Fragment-Based Approach for the Evaluation of NMR Chemical Shifts for Large Biomolecules

Incorporating the Effects of the Solvent Environment. *J. Chem. Theory Comput.* **2017**, *13*, 1147–1158.

(68) Chandy, S. K.; Raghavachari, K. Accurate and Cost-Effective NMR Chemical Shift Predictions for Nucleic Acids Using a Molecules-in-Molecules Fragmentation-Based Method. *J. Chem. Theory Comput.* **2023**, *19*, 544–561.

(69) Jose, K. V. J.; Beckett, D.; Raghavachari, K. Vibrational Circular Dichroism Spectra for Large Molecules through Molecules-in-Molecules Fragment-Based Approach. *J. Chem. Theory Comput.* **2015**, *11*, 4238–4247.

(70) Thapa, B.; Beckett, D.; Jovan Jose, K.; Raghavachari, K. Assessment of fragmentation strategies for large proteins using the multilayer molecules-in-molecules approach. *J. Chem. Theory Comput.* **2018**, *14*, 1383–1394.

(71) 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.

(72) 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.

(73) 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.

(74) 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.

(75) 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.

(76) 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.

(77) 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.

(78) 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.

(79) 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.

(80) Zhang, J. H.; Iyengar, S. S. Graph-IQ(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.

(81) 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.

(82) 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.

(83) Iyengar, S. S.; Saha, D.; Dwivedi, A.; Lopez-Ruiz, M. A.; Kumar, A.; Zhang, J. H.; Ricard, T. C.; Richerme, P.; Sabry, A. Quantum Algorithms for the Study of Electronic Structure and Molecular Dynamics: Novel Computational Protocols. In *Comprehensive Computational Chemistry*; Elsevier, 2023.

(84) Ricard, T. C.; Zhu, X.; Iyengar, S. S. Capturing Weak Interactions in Surface Adsorbate Systems at Coupled Cluster Accuracy: A Graph-Theoretic Molecular Fragmentation Approach Improved through Machine Learning. *J. Chem. Theory Comput.* **2023**, *19*, 8541–8556.

(85) Iyengar, S. S.; Zhang, J. H.; Saha, D.; Ricard, T. C. Graph-IQ(Cl): A Quantum Algorithm with Reduced Quantum Circuit Depth for Electronic Structure. *J. Phys. Chem. A* **2023**, *127*, 9334–9345.

(86) Björklund, A.; Husfeldt, T.; Koivisto, M. Set Partitioning via Inclusion Exclusion. *SIAM J. Comput.* **2009**, *39*, 546–563.

(87) Löwdin, P. O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365–375.

(88) Golub, G. H.; Loan, C. F. V. *Matrix Computations*; The Johns Hopkins University Press: Baltimore, 1996.

(89) Iyengar, S. S.; Frisch, M. J. Effect of Time-Dependent Basis Functions and Their Superposition Error on Atom-Centered Density Matrix Propagation (ADMP): Connections to Wavelet Theory of Multi-Resolution Analysis. *J. Chem. Phys.* **2004**, *121*, 5061–5070.

(90) Dey, T. K.; Shah, N. R. On the number of simplicial complexes in Rd. *Comput. Geom.* **1997**, *8*, 267–277.

(91) Adams, C. C.; Franzosa, R. D. *Introduction to Topology: Pure and Applied*; Pearson Education India, 2008.

(92) Berger, M.; Pansu, P.; Berry, J.-P.; Saint-Raymond, X. Affine spaces. *Problems in Geometry*; Springer, 1984; p 11.

(93) 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.

(94) 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.

(95) 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.

(96) 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.

(97) Kercharoen, T.; Morokuma, K. ONIOM-XS: an extension of the ONIOM method for molecular simulation in condensed phase. *Chem. Phys. Lett.* **2002**, *355*, 257–262.

(98) Huang, L.; Massa, L.; Karle, J. Kernel energy method: Application to DNA. *Biochem* **2005**, *44*, 16747–16752.

(99) 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.

(100) 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.

(101) Gordon, M.; Mullin, J.; Pruitt, S.; Roskop, L.; Slipchenko, L.; Boatz, J. Accurate Methods for Large Molecular Systems. *J. Phys. Chem. B* **2009**, *113*, 9646–9663.

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

(103) Han, J.; Mazack, M. J.; Zhang, P.; Truhlar, D. G.; Gao, J. Quantum mechanical force field for water with explicit electronic polarization. *J. Chem. Phys.* **2013**, *139*, 054503.

(104) Liu, J.; Qi, L.-W.; Zhang, J. Z. H.; He, X. Fragment Quantum Mechanical Method for Large-Sized Ion–Water Clusters. *J. Chem. Theory Comput.* **2017**, *13*, 2021–2034.

(105) Thapa, B.; Raghavachari, K. Energy Decomposition Analysis of Protein-Ligand Interactions Using Molecules-in-Molecules Fragmentation-Based Method. *J. Chem. Inf. Model.* **2019**, *59*, 3474–3484.

(106) Gupta, A. K.; Thapa, B.; Raghavachari, K. Exploring Reaction Energy Profiles Using the Molecules-in-Molecules Fragmentation-Based Approach. *J. Chem. Theory Comput.* **2019**, *15*, 3991–4002.

(107) Xantheas, S. S. *Ab Initio* Studies of Cyclic Water Clusters (H₂O)_N, N = 1–6. II. Analysis of Many-body Interactions. *J. Chem. Phys.* **1994**, *100*, 7523–7534.

- (108) 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.
- (109) 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.
- (110) 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.
- (111) 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.
- (112) 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.
- (113) 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.
- (114) Baer, M. *Beyond Born-Oppenheimer: Conical Intersections and Electronic Nonadiabatic Coupling Terms*; Wiley: New York, 2006.
- (115) Hanna, G.; Kapral, R. Nonadiabatic Dynamics of Condensed Phase Rate Processes. *Acc. Chem. Res.* **2006**, *39*, 21–27.
- (116) Worth, G. A.; Cederbaum, L. S. Beyond Born-Oppenheimer: Molecular Dynamics Through a Conical Intersection. *Annu. Rev. Phys. Chem.* **2004**, *55*, 127–158.
- (117) Jasper, A. W.; Zhu, C.; Nangia, S.; Truhlar, D. G. Introductory Lecture: Nonadiabatic Effects in Chemical Dynamics. *Faraday Discuss.* **2004**, *127*, 1.
- (118) 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.
- (119) 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.
- (120) Yarkony, D. R. Diabolical Conical Intersections. *Rev. Mod. Phys.* **1996**, *68*, 985–1013.
- (121) 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.
- (122) Coker, D. *Computer Simulation in Chemical Physics. Proceedings of the NATO Advanced Study Institute on New Perspectives in Computer Simulation in Chemical Physics*; Springer, 1993; p 315.
- (123) Mecke, K. R.; Wagner, H. Euler characteristic and related measures for random geometric sets. *J. Stat. Phys.* **1991**, *64*, 843–850.
- (124) Rega, N.; Iyengar, S. S.; Voth, G. A.; Schlegel, H. B.; Vreven, T.; Frisch, M. J. Hybrid *Ab-Initio*/Empirical Molecular Dynamics: Combining the ONIOM Scheme with the Atom-Centered Density Matrix Propagation (ADMP) Approach. *J. Phys. Chem. B* **2004**, *108*, 4210–4220.
- (125) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure And/or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384–2393.
- (126) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45*, 1196–1199.
- (127) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- (128) Schlegel, H. B.; Millam, J. M.; Iyengar, S. S.; Voth, G. A.; Daniels, A. D.; Scuseria, G. E.; Frisch, M. J. *Ab initio* molecular dynamics: Propagating the density matrix with Gaussian orbitals. *J. Chem. Phys.* **2001**, *114*, 9758–9763.
- (129) Iyengar, S. S.; Schlegel, H. B.; Millam, J. M. A.; Voth, G.; Scuseria, G. E.; Frisch, M. J. *Ab initio* molecular dynamics: Propagating the density matrix with Gaussian orbitals. II. Generalizations based on mass-weighting, idempotency, energy conservation and choice of initial conditions. *J. Chem. Phys.* **2001**, *115*, 10291–10302.
- (130) Schlegel, H. B.; Iyengar, S. S.; Li, X.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J. *Ab initio* molecular dynamics: Propagating the density matrix with Gaussian orbitals. III. Comparison with Born–Oppenheimer dynamics. *J. Chem. Phys.* **2002**, *117*, 8694–8704.
- (131) Iyengar, S. S.; Schlegel, H. B.; Voth, G. A.; Millam, J. M.; Scuseria, G. E.; Frisch, M. J. *Ab initio* molecular dynamics: Propagating the density matrix with gaussian orbitals. IV. Formal analysis of the deviations from born-oppenheimer dynamics. *Isr. J. Chem.* **2002**, *42*, 191–202.
- (132) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer-Simulation Method for the Calculation of Equilibrium-Constants for the Formation of Physical Clusters of Molecules - Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76*, 637–649.
- (133) Bishop, C. M. *Pattern Recognition and Machine Learning*; Springer: New York, 2006.
- (134) 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.
- (135) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, 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.
- (136) Haykin, S. S. *Neural Networks and Learning Machines*, 3rd ed.; Pearson Education: Upper Saddle River, NJ, 2009.
- (137) Rasmussen, C. E. Gaussian Processes in Machine Learning. In *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2–14, 2003, Tübingen, Germany, August 4–16, 2003*; Bousquet, O., von Luxburg, U., Rätsch, G., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2004; pp 63–71.
- (138) Qu, C.; Yu, Q.; Bowman, J. M. Permutationally Invariant Potential Energy Surfaces. *Annu. Rev. Phys. Chem.* **2018**, *69*, 151–175.
- (139) Nandi, A.; Qu, C.; Bowman, J. M.; et al. 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.
- (140) Li, J.; Varga, Z.; Truhlar, D. G.; Guo, H. Many-Body Permutationally Invariant Polynomial Neural Network Potential Energy Surface for N_4 . *J. Chem. Theory Comput.* **2020**, *16*, 4822–4832.
- (141) Kamath, A.; Vargas-Hernández, R. A.; Krems, R. V.; Carrington, T.; Manzhos, S. Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy. *J. Chem. Phys.* **2018**, *148*, 241702.
- (142) Amabilino, S.; Bratholm, L. A.; Bennie, S. J.; Vaucher, A. C.; Reiher, M.; Glowacki, D. R. Training Neural Nets To Learn Reactive Potential Energy Surfaces Using Interactive Quantum Chemistry in Virtual Reality. *J. Phys. Chem. A* **2019**, *123*, 4486–4499.
- (143) 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.
- (144) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **2021**, *12*, 398.
- (145) Balki, I.; Amirabadi, A.; Levman, J.; Martel, A. L.; Emersic, Z.; Meden, B.; Garcia-Pedrero, A.; Ramirez, S. C.; Kong, D.; Moody, A. R.; Tyrrell, P. N. Sample-Size Determination Methodologies for

Machine Learning in Medical Imaging Research: A Systematic Review. *Can. Assoc. Radiol. J.* **2019**, *70*, 344–353.

(146) Jain, A.; Duin, P.; Mao, J. Statistical pattern recognition: a review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 4–37.

(147) Raudys, S. J.; Jain, A. Small sample size effects in statistical pattern recognition: Recommendations for practitioners and open problems. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *1*, 417–423.

(148) Zhang, W.; Itoh, K.; Tanida, J.; Ichioka, Y. Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Appl. Opt.* **1990**, *29*, 4790–4797.