Article

# Graph-Theory-Based Molecular Fragmentation for Efficient and Accurate Potential Surface Calculations in Multiple Dimensions

Anup Kumar, Nicole DeGregorio, and Srinivasan S. Iyengar*

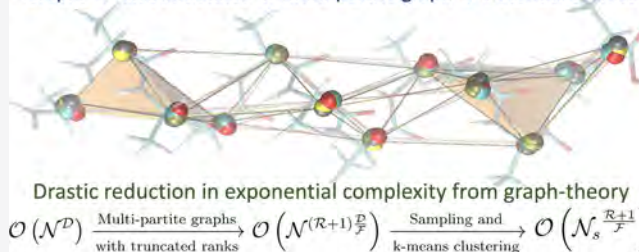ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** We present a multitopology molecular fragmentation approach, based on graph theory, to calculate multidimensional potential energy surfaces in agreement with post-Hartree–Fock levels of theory but at the density functional theory cost. A molecular assembly is coarse-grained into a set of graph-theoretic nodes that are then connected with edges to represent a collection of locally interacting subsystems up to an arbitrary order. Each of the subsystems is treated at two levels of electronic structure theory, the result being used to construct many-body expansions that are embedded within an ONIOM scheme. These expansions converge rapidly with the many-body order (or graphical rank) of subsystems and capture many-body interactions accurately and efficiently. However, multiple graphs, and hence multiple fragmentation topologies, may be defined in molecular configuration space that may arise during conformational sampling or from reactive, bond breaking and bond formation, events. Obtaining the resultant potential surfaces is an exponential scaling proposition, given the number of electronic structure computations needed. We utilize a family of graph-theoretic representations within a variational scheme to obtain multidimensional potential surfaces at a reduced cost. The fast convergence of the graph-theoretic expansion with increasing order of many-body interactions alleviates the exponential scaling cost for computing potential surfaces, with the need to only use molecular fragments that contain a fewer number of quantum nuclear degrees of freedom compared to the full system. This is because the dimensionality of the conformational space sampled by the fragment subsystems is much smaller than the full molecular configurational space. Additionally, we also introduce a multidimensional clustering algorithm, based on physically defined criteria, to reduce the number of energy calculations by orders of magnitude. The molecular systems benchmarked include coupled proton motion in protonated water wires. The potential energy surfaces and multidimensional nuclear eigenstates obtained are shown to be in very good agreement with those from explicit post-Hartree–Fock calculations that become prohibitive as the number of quantum nuclear dimensions grows. The developments here provide a rigorous and efficient alternative to this important chemical physics problem.

Accurate Potential surfaces from graph theoretic fragmentation

$$E_{\{\mathcal{G}_\beta\}}(\mathbf{R}) = E^{level,0}(\mathbf{R}) + \sum_\beta v_\beta(\mathbf{R}) \left( \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in V_r} \Delta E_{\alpha,r,\beta}^{1,0}(\mathbf{R}) \mathcal{M}_{\alpha,r,\beta}^{\mathcal{R}} \right)$$

Graphs → diabatic states → multi-partite graph → Potential Surface

Drastic reduction in exponential complexity from graph-theory

$$\mathcal{O}(\mathcal{N}^{\mathcal{D}}) \xrightarrow[\text{with truncated ranks}]{\text{Multi-partite graphs}} \mathcal{O}\left(\mathcal{N}^{(\mathcal{R}+1)\frac{\mathcal{D}}{\mathcal{F}}}\right) \xrightarrow[\text{k-means clustering}]{\text{Sampling and}} \mathcal{O}\left(\mathcal{N}_s^{\frac{\mathcal{R}+1}{\mathcal{F}}}\right)$$

## 1. INTRODUCTION

The accurate and efficient study of the electronic structure, including electron correlation, and nuclear dynamics is at the core of multiple problems that are critical to materials,[1-3] biological,[4-13] and atmospheric research.[14-17] However, these studies are deeply affected by the steep, polynomial scaling cost of the electronic structure[18-22] and potential exponential scaling of quantum nuclear dynamics.[23-27] The latter is further complicated by the need for accurate potential surfaces on a dimensionally dependent set of the ever-growing number of nuclear geometries.[28-33] With regard to the electronic structure, molecular fragmentation[34-65] and related studies on many-body theory[36,66-86] have recently become an important alternative to traditional studies on electron correlation. The critical goal of these fragmentation methods is to bridge the gap between the quality of the electronic structure and basis sets achievable for small molecular clusters and those used routinely for larger

molecular assemblies. The conventional quality methods have shown a remarkable chemical predictive insight[87-89] scale in a steeply algebraic fashion as the system size grows.[18,21,90-92] Due to these inherent costs, most ab initio molecular dynamics (AIMD) trajectory studies are limited to on-the-fly density functional theory (DFT) calculations with a modest basis, but their use is restricted in many ways.[93-95] Molecular fragmentation-based methods have been noted to remedy the intractability of correlation and extended basis set effects especially for AIMD studies[96-106] and for molecular

potential surfaces.[107,108] Several biological[109−112] studies have also been made possible from the state-of-the-art molecular fragmentation methods.

In a series of publications,[99,102−104,106,107,113−115] we have shown how graph-theoretic methods may be used to adaptively construct many-body approximations within ONIOM[39,116−120] to obtain accurate post-Hartree−Fock potential surfaces[107,114] and AIMD trajectories[99,102−104,106,113] at a reduced computational cost. Beginning from an ONIOM-type decomposition of the system, the individual extrapolatory components within ONIOM (the so-called "model" and "real" components of the calculation) are obtained using many-body theory,[36,66−86] which is written in a general and adaptive fashion using graph theory. The idea of using a graph-theoretic description to construct many-body approximations allows the use of multiple powerful libraries[121,122] in Python, which are extremely well refined and robust and allow for implementation of a completely general partitioning or molecular fragmentation scheme. As a result, the many-body decomposition used here, within ONIOM, and in refs 103, 104, 106, 107, 114, and 115 is adaptive and system-independent. Although our formalism in refs 99, 102−104, 106, 107, and 113−115 is derived from ONIOM,[39,116−120] it also has close connections to other fragmentation and many-body methods including the multi-centered QM/QM formalism,[35,36] the molecular tailoring approach,[37,38] the ONIOM-XO method,[39] the hybrid many-body interaction (HMBI) model,[44,123] the molecules-in-molecules (MIM) methodology,[40−43] and the incremental method,[44−48] to name just a few. Indeed, there are several other fragmentation methods[49−65] available, but the approaches in refs 35, 38−42, 99, 102, 103, and 113 include long-range electronic effects through a full-system low-level calculation, much in the same vein as the ONIOM[39,116−120] method.

The use of accurate analytical gradients associated with the energy expression in refs 99, 102−104, 106, 107, and 113−115 yields both extended Lagrangian-[102,113] and Born-Oppenheimer[99,102,113]-based AIMD simulations that can be performed at accuracy comparable to CCSD and MP2 levels of theory with DFT-computational expense. We have also shown that weak interactions (specifically hydrogen bonds) can be accurately captured and efficient approximations to large-basis AIMD trajectories can be constructed through computational effort commensurate with much smaller basis set sizes. For example, in ref 104, AIMD trajectories are constructed in agreement with basis sets such as 6-311+ +G(2df,2pd) with computational effort commensurate with those from smaller basis sets such as 6-31+G(d), for polypeptide systems with 100+ atoms. This is carried out using both extended Lagrangian and Born-Oppenheimer techniques. Furthermore, we have also shown in ref 115 that using this graph-theoretic technique, condensed-phase simulations on surfaces and bulk systems may be constructed at hybrid DFT (Rung-4 functionals) accuracy using gradient-corrected DFT (Rung-2 functionals) effort. This has been shown for both atom-centered and plane-wave basis sets. Furthermore, as seen in ref 115, this methodology allows one to supplement plane-wave-based DFT calculations, with local, heterogeneous, electronic effects, as these may arise from substrate adsorption, using atom-centered basis functions. In this paper, we generalize our formalism for graph-theoretic interpolation of potential energy surfaces to allow a completely automated procedure that works efficiently for arbitrary dimensional problems. The main features of the method discussed here are as follows: as noted in ref 107, the graphical decomposition is not unique when a family of molecular configurations are considered as needed during a potential surface calculation. For example, when excess protons or acidic groups are locally present, as they might be in water clusters and in solvated/hydrogen-bonded amino acid groups, proton transfer changes the nature of the coarse-grained graphical units (or nodes in the graph) and also changes the graphical connectivity. Similarly, large-scale rearrangement of molecular frameworks, including rearrangement of solvation shells and general conformational dynamics, also allows for a change in the graph-theoretic adjacency description. In all such cases, multiple graphical structures must be considered *simultaneously*. In this regard, the key idea in ref 107 is to interpret the potential surfaces obtained from each of the graphs as individual diabatic states[124−134] and obtain the overall potential surface through a linear combination of such diabatic states, constructed using a variational procedure, as is carried out within many valence bond-based methods.[127,135−147] This was shown to work very well in ref 107 for a few degrees of freedom. In Section 2, we generalize this procedure to multiple dimensions.

The paper is organized as follows: in Section 2, we first present a brief survey of our graph-theoretic approach for the low-cost electronic structure, AIMD, and potential surface calculations, followed by the generalization of the same to obtain multidimensional potential surfaces obtained from combining multiple graphical representations, or multiple fragmentational depictions, of a given molecular assembly. Computational aspects of the graph-theoretic approach are discussed in Section 3 with more details provided in the Supporting Information. The study of potential energy surfaces in a protonated water-wire system is provided in Section 4. Additionally, Appendices A through D complement the discussion in Sections 2 through 4. Conclusions are given in Section 5.

## 2. MULTITOPOLOGY MOLECULAR FRAGMENTATION PROCEDURE FOR EFFICIENT AND ACCURATE POTENTIAL SURFACE EVALUATIONS

The basic idea behind the approach in refs 99, 102−104, 106, 107, and 113−115 is as follows: A molecular assembly is first divided into a set of "nodes" or "vertices". In refs 99, 102−104, 106, 107, and 113−115, as part of the study on protonated water clusters and polypeptide fragments and the study of adsorption of organic impurities on the surface of water, these "coarse-grained" units are considered to be single-amino-acid groups within a polypeptide sequence, adsorbed organic molecules such as methanol and methane molecules on the surface water that is treated using periodic boundary conditions, water molecules, or their protonated form. However, the implementation is general enough to include other nodal definitions. Following this, based on a chosen two-body interaction envelope, edges are defined, leading to a map between a specific molecular geometry defined by the nuclear coordinates, $\mathbf{R}$, and a graph, $\mathcal{G}$, that is defined using the set of vertices and edges, that is, $\mathcal{G} \equiv \{\mathbf{V}_0; \mathbf{V}_1\}$. Here, $\mathbf{V}_0$ is the set of vertices or coarse-grained partitions in the molecular assembly and $\mathbf{V}_1$ is the set of edges that capture interactions between these vertices or

molecular fragments. However, once such a graph is defined, higher rank objects up to a maximum rank $\mathcal{R}$ are also specified by the graph, as given by

$$\{\mathbf{V_r} | r = 0, ..., \mathcal{R}\} \equiv \{\mathbf{V_0}, \mathbf{V_1}, \mathbf{V_2}, ..., \mathbf{V_r}, ..., \mathbf{V_\mathcal{R}}\} \tag{1}$$

The quantity $\mathbf{V_r}$ represents the set of all rank-$r$ entities in the graph that may be used to capture the interactions between $(r + 1)$ nodes or monomers as we carry out below in eqs 2 and 3. We further include within $\mathbf{V_r}$ only those rank-$r$ entities that are fully connected. Thus, a triangle is included in $\mathbf{V_2}$ only if all nodes inside the triangle are already connected by edges and are included in $\mathbf{V_1}$. Similarly, a specific rank-3 object (a tetrahedron) is included in $\mathbf{V_3}$, only if all triangles, or trimers, within the rank-4 object are also included within $\mathbf{V_2}$. Such rank-$r$ objects, where all pairs of nodes are completely connected, are called simplexes.[148−152] This geometrical consideration is critical in our approach, as we will see, and leads to a significant connection to many-body expansions in the next paragraph.

One such example of a graphical representation of a molecular system is depicted in Figure 1 where an amino acid
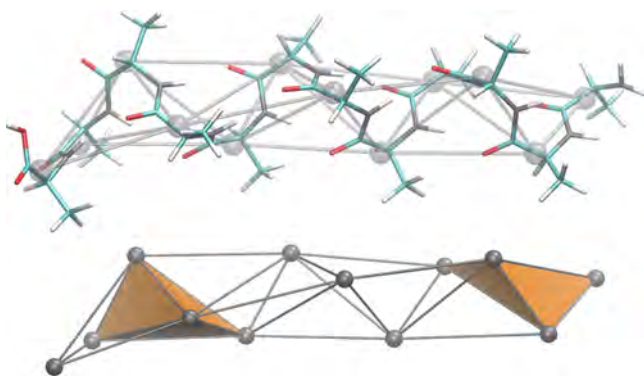


**Figure 1.** Molecular graph representation (shown at the bottom) of a polypeptide sequence (shown on the top) where each amino-acid group is treated as a node. Higher order simplexes are described as per eq 1, and two simplexes of rank-3 (i.e., tetrahedrons) are highlighted using orange shading.

chain containing 12 alanine units is coarse-grained into a molecular graph. For the molecular graph thus-obtained, two of the simplexes[148−152] in $\mathbf{V_3}$ are highlighted in the lower panel of Figure 1 as orange shaded regions. The fact that these simplexes are well-separated, that is, the orange simplexes do not spatially overlap, has a critical role in the efficiency of the algorithm presented in this paper.

Once a system is graphically partitioned as described above, many-body expansions are introduced to obtain

$$E_{\text{MBE},\mathcal{R}}^{\text{level,I}}(\mathbf{R}) = \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V_r}} E_{\alpha,r}^{\text{level,I}}(\mathbf{R}) \left[ \sum_{m=r}^{\mathcal{R}} (-1)^m p_\alpha^{r,m} \right]$$

$$= \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V_r}} E_{\alpha,r}^{\text{level,I}}(\mathbf{R}) \mathcal{M}_{\alpha,r}^{\mathcal{R}} \tag{2}$$

where "level, I" represents the level of electronic structure theory used to construct this expansion, $p_\alpha^{r,m}$ is the number of times the $\alpha$th $(r + 1)$-body term (in set $\mathbf{V_r}$) appears in all $(m + 1)$-body terms (in set $\mathbf{V_m}$) for $m \geq r$, and, consequently,

$\mathcal{M}_{\alpha,r}^{\mathcal{R}}$ is the number of times the $\alpha$th $(r + 1)$-body term appears in all objects of rank greater than or equal to $r$, until the maximum rank, $\mathcal{R}$. Thus, eq 2 is essentially the standard many-body expansion[36,66−86] written and implemented using graph-theoretic methods. This fact is explicitly shown in ref 115 and also here in Appendix A. A set of two such many-body approximations, constructed at two different levels of theory labeled "level, 1" and "level, 0", are then combined using an ONIOM scheme to obtain

$$E_{\mathcal{R}}^{1,0}(\mathbf{R}) = E^{\text{level,0}}(\mathbf{R}) + E_{\text{MBE},\mathcal{R}}^{\text{level,1}} - E_{\text{MBE},\mathcal{R}}^{\text{level,0}}$$

$$= E^{\text{level,0}}(\mathbf{R}) + \sum_{r=0}^{\mathcal{R}} (-1)^r$$
$$\sum_{\alpha \in \mathbf{V_r}} (E_{\alpha,r}^{\text{level,1}}(\mathbf{R}) - E_{\alpha,r}^{\text{level,0}}(\mathbf{R})) \mathcal{M}_{\alpha,r}^{\mathcal{R}}$$

$$= E^{\text{level,0}} + \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in V_r} \Delta E_{\alpha,r}^{1,0} \mathcal{M}_{\alpha,r}^{\mathcal{R}}$$

$$= E^{\text{level,0}} + \Delta \text{MBE}_{\mathcal{R}}^{\text{level,1;level,0}} \tag{3}$$

where

$$\Delta E_{\alpha,r}^{1,0} \equiv E_{\alpha,r}^{\text{level,1}}(\mathbf{R}) - E_{\alpha,r}^{\text{level,0}}(\mathbf{R}) \tag{4}$$

Furthermore

$$\Delta \text{MBE}_{\mathcal{R}}^{\text{level,1;level,0}} \equiv E_{\text{MBE},\mathcal{R}}^{\text{level,1}} - E_{\text{MBE},\mathcal{R}}^{\text{level,0}}$$

$$= \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V_r}} \Delta E_{\alpha,r}^{1,0} \mathcal{M}_{\alpha,r}^{\mathcal{R}} \tag{5}$$

is the correction term added to $E^{\text{level,0}}$, in eq 3, that includes the additional many-body contributions that arise from "level, 1" treatment, as compared to "level, 0" treatment. As is standard in ONIOM, the first term on the right side of eq 3, $E^{\text{level,0}}$, represents the lower-level approximation to the full system, which is then improved upon through the difference $(E_{\text{MBE},\mathcal{R}}^{\text{level,1}} - E_{\text{MBE},\mathcal{R}}^{\text{level,0}})$. Additionally, the $\mathbf{R}$ dependence of $\alpha$ on the third and fourth lines on eq 3 is dropped for the purpose of clarity. In refs 99, 102−104, 106, 107, and 113−115, eq 3 is implemented using two levels of theory such as DFT (for level, 0) and post-Hartree−Fock methods (MP2 and CCSD for level, 1) or similarly low- and high-level Gaussian basis sets in ref 104 to obtain accurate and efficient AIMD trajectories. In refs 106 and 115, we have demonstrated the rapidly convergent nature of eq 3 and contrast this behavior with standard many-body theory as derived from the direct application of eq 2. Furthermore, in refs 106 and 115, we have also shown that eq 3 is numerically convergent at order $\mathcal{R} = 2$ or less (three-body terms) for water clusters, aqueous organic-solute/solvent interfaces, and bulk water systems. (See Appendix B.) Thus, through refs 99, 102−104, and 113−115, we have established an automated approach, that has connections to ONIOM, molecular fragmentation, and many-body theory, to carry out (a) post-Hartree−Fock AIMD at the DFT cost,[99,102,113] (b) large basis set AIMD at low basis costs,[104] (c) condensed-phase studies with Rung-4 (hybrid) functional accuracy but at a Rung-2 (pure) functional computational cost,[115] and (d) plane-wave-based condensed-phase calculations that are suitably augmented

with localized, polarized, and diffuse Gaussian-type basis functions but at little additional cost.[115]

## 2.1. Generalizations to Eq 3: Multiple Simultaneous Graphical Depictions and Fragmentation Schemes to Obtain Potential Energy Surfaces for Quantum Nuclear Treatment.

The goal of the current paper is to generalize the ideas introduced in refs 106 and 107 to achieve an efficient evaluation of multidimensional potential energy surface calculations. To begin, it is recognized in refs 106 and 107 that the graphs obtained by mapping nuclear geometry, that is, $\mathbf{R} \rightarrow \mathcal{G}$, encode *all* the necessary interactions involved in a system, and hence, the vertices, edges, faces, and higher-order simplexes of such a graph may change during potential surface evaluations[107] and during AIMD trajectories.[106] For example, the way in which the vertices are connected may change, due to new hydrogen bonds, or the vertices themselves may change perhaps through the addition or removal of protons, or more generally as part of other reactive processes. In general, one may obtain a family of graphs $\{\mathcal{G}_\beta\}$ each providing a different representation of energy and gradients, that is

$$\mathcal{G}_\beta \Rightarrow E_{\mathcal{R},\beta}^{1,0}(\mathbf{R}) \tag{6}$$

The key idea in refs 106 and 107 is to write the global potential surface as a linear combination of such a family of surfaces, and thus, using eq 3, we obtain

$$E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\mathbf{R}) = \sum_\beta v_\beta(\mathbf{R}) E_{\mathcal{R},\beta}^{1,0}(\mathbf{R}) \tag{7}$$

where the notation $E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\mathbf{R})$, used throughout this paper, is meant to imply that a family of graphs $\{\mathcal{G}_\beta\}$ are used to compute the potential energy surface. Using eq 3, we obtain

$$E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\mathbf{R}) = \sum_\beta v_\beta(\mathbf{R})[E^{level,0}(\mathbf{R}) + E_{MBE,\mathcal{R},\beta}^{level,1}(\mathbf{R}) - E_{MBE,\mathcal{R},\beta}^{level,0}(\mathbf{R})]$$

$$= \left(\sum_\beta v_\beta(\mathbf{R})\right) E^{level,0}(\mathbf{R})$$

$$+ \sum_\beta v_\beta(\mathbf{R}) \left(\sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_r^\beta} \Delta E_{\alpha,r,\beta}^{1,0}(\mathbf{R}) \mathcal{M}_{\alpha,r,\beta}^{\mathcal{R}}\right) \tag{8}$$

where the fact that each graph may have a different set of nodes and edges is depicted using the notation $\mathcal{G}_\beta \equiv \{\mathbf{V}_0^\beta; \mathbf{V}_1^\beta\}$, and thus, $\mathbf{V}_r^\beta$ is a set of all *r*-rank simplexes in $\mathcal{G}_\beta$. The probabilistic coefficients, $\{v_\beta\}$, in eq 8 are obtained in ref 107 by (a) interpreting the surfaces, $\{E_{\mathcal{R},\beta}^{1,0}(\mathbf{R})\}$, as crude diabatic states[124-134] and (b) introducing a variational procedure to obtain $\{v_\beta\}$. In this sense, we remain strongly influenced by the work in the valence-bond literature.[127,135-147] Therefore, in essence, we interpret each fragmentation protocol (or each graphical connectivity that leads to a fragmentation protocol) as yielding a valence bond-type diabatic state, hence leading to a family of diabatic potential surfaces $\{E_{\mathcal{R},\beta}^{1,0}\}$ in eq 7. The coefficients $v_\beta(\mathbf{R})$ then yield adiabatic states as a linear combination of the graphically determined molecular fragmentation schemes.

This paper discusses two main computational issues that arise from the overarching goal of computing eq 8 (and eq 9 below) for arbitrary nuclear dimensions. One aspect deals

with computing the diabatic potential surfaces $\{E_{\mathcal{R},\beta}^{1,0}\}$ in eq 7, using the individual graph-theoretic fragmentation protocols. The second aspect deals with computing the adiabatic state coefficients, $\{v_\beta(\mathbf{R})\}$. Furthermore, the summation $(\sum_\beta v_\beta(\mathbf{R}))$ in eq 8 is the identity, and hence, eq 8 reduces to

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

$$= E^{level,0}(\mathbf{R}) + \sum_\beta v_\beta(\mathbf{R})(\Delta MBE_{\mathcal{R},\beta}^{level,1;level,0}) \tag{9}$$

where we have designated the correction term in eq 5 using the graph index $\beta$ in the subscript, that is, $\Delta MBE_{\mathcal{R},\beta}^{level,1;level,0}$. As noted above, in the remaining portion of this paper, based on eq 9, we develop a scheme to compute multidimensional potentials efficiently, at post-Hartree−Fock accuracy, with the DFT cost. The associated potential surfaces are then used to compute quantum nuclear eigenstates that help gauge the error in these potential surfaces.

## 2.2. Local Nature of the Expression in Eq 9.

The left-hand side of eq 9 provides energy for a molecular system with nuclear coordinates given by $\mathbf{R}$. The right-hand side in the equation is a combination of the full-system low-level potential energy $E^{level,0}(\mathbf{R})$ and an $v_\beta(\mathbf{R})$-weighted sum of energy-correction terms ($\Delta MBE_{\mathcal{R},\beta}^{level,1;level,0}$, within parentheses) arising from each graphical description, $\mathcal{G}_\beta$. Essentially, what we have carried out here is to improve the energy of a full-system low-level calculation, $E^{level,0}(\mathbf{R})$, at an additional expense of high- and low-level energy calculations of molecular fragments (second term in eq 9), that are obtained using a family of graphs. While the goal of eq 3, as discussed in refs 99, 102−104, 106, 107, and 113−115, has been to achieve agreement with $E^{level,1}(\mathbf{R})$, which is the full system energy at level, 1 theory, that may have a steep algebraic computational scaling (see Appendix B), the goal of eq 9 is to achieve a similar level, 1 quality potential energy surface, but at a reduced cost as implied from eq 9. (See Section 3 and Appendix C for a detailed discussion on computational scaling.)

For each energy-correction term in eq 9, the key $\mathcal{G}_\beta$-dependent portions are the energy differences, $\Delta E_{\alpha,r,\beta}^{1,0}(\mathbf{R})$, and the multiplicity-factor, $\mathcal{M}_{\alpha,r,\beta}^{\mathcal{R}}$, included to avoid overcounting as in the energy correction as in eq 3. Both of these terms, through dependence on $\alpha \in \mathbf{V}_r^\beta$, only contain the molecular fragments with $(r + 1)$ nodes and hence do not depend on the complete set of molecular coordinates that span the full dimensionality of $\mathbf{R}$. As the energy-correction term only requires the energy calculations for the molecular fragments and not on the full system, the number of nuclear degrees of freedom involved in the fragments that determine $\{\Delta E_{\alpha,r,\beta}^{1,0}(\mathbf{R})\}$ is, in general, drastically lower than the full dimensionality of $\mathbf{R}$. For example, let us consider the molecular system coarse-grained in Figures 1 and 2. In Figure 2, the nodes and edges for different graphical decompositions are depicted using different colors. As a
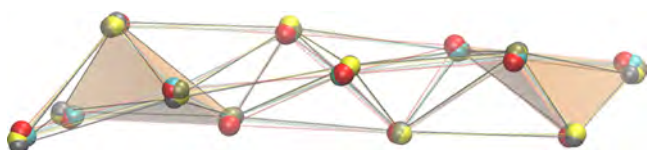
**Figure 2.** Graphical representation of the coupled quantum nuclear and electronic description of the system shown in Figure 1. This figure complements eq 9 and the associated discussion. The nodes in different colors represent nuclear delocalization due to quantum effects. In this paper and in refs 99, 102−104, 106, 107, and 113−115, the electronic structure is simplified using graph theory. Here, the simplexes shown using orange shadings contain only a small subset of all coupled quantum nuclear dimensions. As a result, the corresponding $\Delta E_{\alpha,r,\beta}^{1,0}(\mathbf{R})$ terms in eq 9 only depend on a subset of the quantum nuclear dimensions, which enormously reduces computational complexity.

result, the rank-3 simplexes (tetrahedrons) and all other simplexes in Figure 2 contain far fewer nuclear degrees of freedom as compared to the full system in Figure 2. In addition, the nuclear degrees of freedom captured within the orange-colored simplex shown on the left in Figure 2 are, in general, different from those sampled by the other simplexes such as the orange-colored simplex on the right. This allows us to write the nuclear potential energy surface in terms of the potential energy functions corresponding to the simplexes. Here, the family of electronic structure calculations needed in eq 9 is

$$E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\mathbf{R}) \leftarrow \{E^{\text{level},0}(\mathbf{R}), \{E_{\alpha,r,\beta}^{\text{level},1}(\mathbf{R}_{\alpha,r}), E_{\alpha,r,\beta}^{\text{level},0}(\mathbf{R}_{\alpha,r})\}\} \tag{10}$$

and we have specifically noted the fact that the molecular fragments obtained from graph theory do not sample the full molecular space; the corresponding nuclear dimensions sampled are represented above using the symbol $\mathbf{R}_{\alpha,r}$.

This idea of dimensional reduction obtained from eq 9 is similar to the high-dimensional model representation,[29,153] weighted sum of products of approximations,[154,155] and the permutationally invariant polynomials[156] approach; influenced by the early work of Kolmogorov[157,158] and Sobol',[159] a high-dimensional function is written as a linear combination of lower-dimensional functions, where these lower-dimensional functions may be obtained from random sampling[28,153] or using neural networks.[29,160] To clarify the connection to these dimensional reduction techniques, we emphasize that, in eq 9

$$\Delta\text{MBE}_{\mathcal{R},\beta}^{\text{level},1;\text{level},0} = \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_r^\beta} \Delta E_{\alpha,r,\beta}^{1,0}(\mathbf{R}) \, \mathcal{M}_{\alpha,r,\beta}^{\mathcal{R}} \tag{11}$$

where the sum over "$r$" is essentially a sum over lower-dimensional functions that are only dependent on, at the most, "$r$" degrees of freedom. This, of course, is also the essence of many-body theory[36,66−86] Similar ideas have also been explored in ref 161, where methods that utilized expressions similar to eq 2 are used to compute potential energy correction surfaces that are similar to those presented
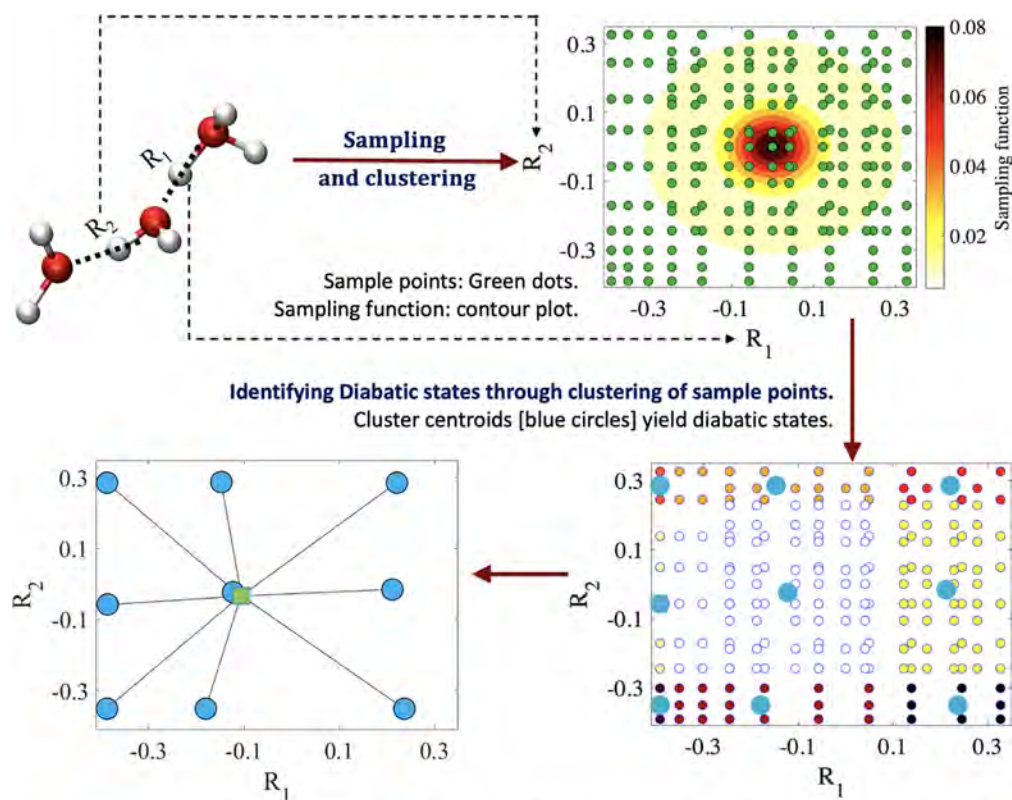


**Figure 3.** Essential features of the algorithm described in Section 3 and in the Supporting Information.

in eq 10. However, the approach here differs in that eq 9 includes (a) all ranks up to a given order $\mathcal{R}$, (b) multiple levels of theory within an ONIOM format, and (c) a recipe to compute post-Hartree−Fock surfaces, potentially at the expense of level, 0 theory, as has been discussed in ref 103 and in Section 3 and Appendix B and Appendix C. These connections are also made more explicit in Appendix A. Thus, considering that the simplexes indexed by $(\alpha,r)$ partition the Euclidean coordinate space of $\mathbf{R}$, each contribution, $\Delta E_{\alpha,r,\beta}^{1,0}(\mathbf{R})$, to the summation over $(\alpha,r)$ in eq 9, only needs to be computed inside a portion of the full molecular configurational space (see eq 10), thus contributing enormously to the computational efficiency.

## 3. COMPUTATIONAL ASPECTS REGARDING IMPLEMENTATION OF EQ 9

We wish to provide general prescriptions regarding the implementation of eq 9 toward constructing multidimensional potential surfaces. There are multiple considerations to this implementation. The resultant algorithm is illustrated through Figure 3 using protonated water clusters as examples, where we intend to study the coupled quantum behavior of multiple shared protons. For example, in Figure 3, we display a protonated water trimer along with the real space span of the two shared protons (named $R_1$ and $R_2$ in Figure 3). These and the other associated parts of Figure 3 are used to explain the algorithm here to obtain potential surfaces in arbitrary dimensions. The computational tests are carried out on an extended version of this system, where multiple protons are treated in a quantum-mechanical fashion to allow the coupled correlated treatment of such systems. The Supporting Information sections complement the discussion in this section.

While the fragment energy components, $\{E_{\alpha,r,\beta}^{\text{level},1}(\mathbf{R}_{\alpha,r}),$ $E_{\alpha,r,\beta}^{\text{level},0}(\mathbf{R}_{\alpha,r})\}$ in eqs 9 and 10, are only dependent on a smaller portion of the full configurational space, as outlined in Section 2.2 and Appendix C, the quantity $E^{\text{level},0}(\mathbf{R})$ depends on the full configurational space. Thus, we first provide a sampling measure and the associated clustering algorithm that allows us to choose a representative set of sample points in multidimensional space where the individual terms in eqs 9 and 10 are to be evaluated. The sampling measure introduced here is discussed in more detail in Section S2 and is influenced by previous work,[32,162−164] where we have introduced Shannon entropy-based sampling measures[163,164] along with measures that depend on local length scales of potentials[32,162] to help determine a subset of sample points that could be truly representative of a potential surface. The studies in refs 32 and 162−166 have also demonstrated the computational efficiency and accuracy of such measures. However, here, our sampling measure is simpler and less physically motivated as compared to those in refs 32 and 162−164; here, it is tailored toward the applications in Section 4. More precisely, while the implementations in refs 32 and 162−164 allow for a nondirect product grid of sample points, the implementation here only uses a direct product of sample points. Details regarding the sampling measure are provided in Section S2. As stated earlier, however, all essential features of the algorithm are captured in Figure 3. Using the algorithm in Section S2, specifically the sampling function in eq S2 (which is shown as a contour plot, on the top-right of Figure 3), we sample the two-dimensional potential energy space using a

weighted $k$-means-clustering approach[167−170] which works in a similar fashion to the centroid Voronoi formalism[171−175] and uses Lloyd's algorithm.[176] For example, the illustration on the top-right of Figure 3 depicts the location for 200 sample points chosen from a regular grid containing approximately 10,000 points (i.e., 2% of the total number of grid points). These samples are regions where electronic structure calculations $\{E^{\text{level},0}(\mathbf{R}); \{E_{\alpha,r,\beta}^{\text{level},1}(\mathbf{R}),E_{\alpha,r,\beta}^{\text{level},0}(\mathbf{R})\}\}$ in eq 10 are to be performed. The sample points are shown using green circles on the top-right and using different colors on the bottom-right of Figure 3 for reasons that will soon become clear. Detailed discussions on the algorithmic aspects are provided in Section S2.

The sampling measures yield a family of sample points in the multidimensional potential surface space. However, multiple sample points obtained from our sampling measure (see Section S2) may have similar graphical connectivities. As a result, a clustering algorithm is discussed in Section S3, where the sample points obtained are clustered together based on the similarity of their respective graphical definitions. The associated cluster centroids are shown using blue circles (bottom figures) in Figure 3. The clustering algorithm provides a representative family of graphs: $\{\mathcal{G}_\beta\}$. For every given graph within this set, $\{\mathcal{G}_\beta\}$, the electronic energy is computed at all the sample points (green circles on the top-right of Figure 3) obtained from our sampling treatment, suitable interpolation of which yields crude diabatic states, $\{E_{\mathcal{R},\beta}^{1,0}(\mathbf{R})\}$. A key step in our analysis is the interpretation of $\{E_{\mathcal{R},\beta}^{1,0}(\mathbf{R})\}$, in eq 8, as crude diabatic states.[124−134] Clearly, the presence of nine such diabatic states (nine blue circles inside each of the two subfigures in the lower panel of Figure 3) for the two degrees of freedom is realized through the natural interpretation that each proton has three Marcus-type diabatic locations, one on the donor side, one on the acceptor side, and a third that is roughly shared between the donor and acceptor sites that may be critical for short-strong hydrogen bonds.[177−181] This yields the nine diabatic states from the nine graphical connectivities and produces nine potential surfaces for the illustrated two-dimensional case. In higher dimensions, this potentially leads to a maximum of $3^{N_{\text{Dim}}}$ such states. However, as we see in Section 4, the number of accessible diabatic states is far fewer as compared to $3^{N_{\text{Dim}}}$. This is due to the fact that for any given set of donor−acceptor distances, only a subset of the $3^{N_{\text{Dim}}}$ diabatic states are energetically accessible.

As noted earlier, in computing the weights, $v_\beta(\mathbf{R})$, we remain influenced by the empirical valence bond (EVB) formalism.[127,135−147] Thus, each graphical description $\mathcal{G}_\beta$ depicts a valence bond-type diabatic state, and hence, the coefficients $v_\beta(\mathbf{R})$ transform from the family of "diabatic states", $\{E_{\mathcal{R},\beta}^{1,0}\}$, to the "adiabatic states" depicted on the left of eq 9. Section S5 is devoted to summarizing the Hamiltonian formalism adapted from EVB for our purposes here and is also discussed in detail in ref 107. Thus, the adiabatic linear combinations yield

$$\{E_{\mathcal{R},\beta}^{1,0}(\mathbf{R})\} \xrightarrow{v_\beta(\mathbf{R})} E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\mathbf{R}) \tag{12}$$

through the variational procedure that is designed in a fashion commensurate with EVB[127,135−147] theory. The procedure is also illustrated in Figure 3, where the molecular

**Table 1. Grid Parameters for Potential Energy Surface Calculations**[a]



| dimensionality | grid dimensions | total number of grid points[b] | number of sampling points[c] |
|---|---|---|---|
| 2 $(R_1, R_2)$ | 0.8 Å × 0.8 Å | $99^2 = 9801$ | 98 (1%)−2048 (20%)[d] |
| 3 $(R_1, R_2, R_3)$ | 0.8 Å × 0.8 Å × 0.8Å | $49^3 = 117{,}649$ | 3456 (3%) |
| 4 $(R_1, R_2, R_3, R_4)$ | 0.8 Å × 0.8 Å × 0.8 Å × 0.8 Å | $49^4 = 5{,}764{,}801$ | 41,472 (1%) |
| 5 $(R_1, R_2, R_3, R_4, R_5)$ | 0.8 Å × 0.8 Å × 0.8 Å | $23^5 = 6{,}436{,}343$ | 65,536 (1%) |
| | 0.8 Å × 0.8 Å | | |

[a]The actual physical dimensions are along the proton transfer coordinates of a water-wire molecule $((H_2O)_6H^+)$ shown in the figure immediately above the table. The variables $R_1, R_2, ..., R_5$ are represented using discretized one-dimensional grids aligned along the donor−acceptor axes noted in the figure using dashed lines. The donor−acceptor distances and the shared proton's distances from the donor are also shown for all the grids. In the Supporting Information, we provide a brief section further describing the grid. [b]This total number also represents the number of electronic structure energies that may be necessary and depicts the complexity in the absence of sampling. [c]Actual number of electronic energies computed when the sampling function, discussed in Section 4 and in the Supporting Information, is used. The electronic energies are obtained using eq 9. [d]Range of sample-point densities considered.

**Table 2. Error Estimates and Grid Compression Factors (from Sampling) for the Two-Dimensional Case**[a]

| | 0.01[b] | 0.02[b] | 0.03[b] | 0.04[b] | 0.05[b] | 0.06[b] | 0.07[b] | 0.08[b] | 0.09[b] | 0.10[b] | 0.21[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\Delta\epsilon_{V,\omega=\rho_0}$[c] | 0.23 | 0.11 | 0.11 | 0.13 | 0.12 | 0.16 | 0.18 | 0.11 | 0.11 | 0.11 | 0.11 |
| $\Delta\epsilon_{V,\omega=B}$[c] | 0.07 | 0.06 | 0.05 | 0.04 | 0.04 | 0.04 | 0.05 | 0.04 | 0.03 | 0.04 | 0.05 |
| $\Delta\epsilon_{E\alpha=0}$[d] | 0.22 | 0.03 | 0.06 | 0.12 | 0.08 | 0.16 | 0.18 | 0.07 | 0.08 | 0.07 | 0.06 |

[a]The number of sample points dictates the number of electronic structure energies needed and hence depicts the complexity. Also, see Table 1. [b]Ratio of the number of sampling points (last column in Table 1) to the total number of grid points (penultimate column in Table 1). The numbers along the first row depict the extent of grid compression due to sampling. [c]Potential surface error estimates (in kcal/mol) as per eq D1. The errors $\Delta\epsilon_{V,\omega=\rho_0}$ and $\Delta\epsilon_{V,\omega=B}$ are obtained using the weighting function $\omega$ as $\rho_0$, the ground-state wave packet density, and the Boltzmann factor with a temperature of 300 K. See Appendices D and E. [d]Eigenvalue error estimates (in kcal/mol) as per eq D2. Ground-state, first-excited-state, and second-excited-state eigenvalues are 6.44, 12.05, and 13.39 kcal/mol, respectively, above the minimum. See Appendices D and E.

graph centroids (blue spheres) yield diabatic states, $\{E_{\mathcal{R},\beta}^{1,0}(\mathbf{R})\}$, that are utilized with the variational formalism in Section S5 (also see ref 107 for a detailed description of this algorithm) to obtain the adiabatic state, $E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\mathbf{R})$.

## 4. EFFICIENCY AND ACCURACY OF THE MULTITOPLOGY FORMALISM AND MULTIDIMENSIONAL POTENTIAL SURFACES

We examine the accuracy and efficiency of the formalism discussed above in obtaining multidimensional potential energy surfaces at post-Hartree−Fock levels of theory at reduced computational effort. The precise reduction in costs is summarized in Appendix C. The chemical system considered is a protonated water wire[4−7,107,145] $((H_2O)_6H^+)$ shown in the top panel of Table 1. This class of problems is prevalent in a wide range of biological systems including membranes,[182−184] ion channels[185] where it supports charge transport, fuel cell technologies,[186] and photosynthetic reaction active sites.[183] Proton transfer in these systems follows the Grotthuss mechanism where a concatenated sequence of water molecules makes the transport of excess charge efficient compared to mass transport required for other ions.[4] Furthermore, the light nature of the transferring proton makes quantum nuclear effects critical,[127,187−189] and specifically, coupled multidimensional quantum-nuclear ef-

fects to be important. Hence, we use this system as a test bed for our multidimensional potential energy surface calculations.

In Appendix D, we introduce a series of error estimates to gauge the accuracy of the potentials obtained from the formalism discussed above. For all the graph-theoretic potential energy calculations performed here, we have used B3LYP/6-31++g(d,p) and MP2/6-31++g(d,p) for "level, 0" and "level, 1" (from eqs 8 and 4), respectively. That is, our target multidimensional potential surface is the MP2/6-31++g(d,p) level of theory. For computing errors, the reference level of theory is always the full system (without fragmentation) depicted as "level, 1" (i.e., MP2/6-31++g(d,p)). In Section 4.1, we address the issue of sampling multidimensional potentials, as necessitated by the coupled treatment of multiple protons within the water-wire problem considered here. As noted in eqs C1, C2, C3, and C4, this is complicated by the intrinsically steep scaling of electronic structure calculations needed during potential surface calculations, as well as the number of these calculations. These challenges are substantially alleviated by the use of eq 9 which provides an approach to reduce the number of quantum nuclear dimensions that the fragment energies span and through the sampling function discussed in Section S2. We also analyze how these sampled grid points are distributed. Following Section 4.1, we evaluate the accuracy

and efficiency of the methods introduced in Section 2 in computing higher dimensional potential surfaces.

**4.1. Analysis of the Sampling Measure and Its Effect on Accurate Multidimensional Potentials When Combined with the Graph-Theoretic Protocol.** We consider a range of sample sets as outlined in Tables 1 and 2. Specifically, the donor−acceptor span of the five hydrogen bonds on the water wire is treated as a single, coupled, quantum mechanical problem, and by treating them in a sequential fashion (rows in Table 1), we allow the complexity of the problem to be increased by one shared proton dimension at a time. This allows for the concerted, correlated treatment of the Grotthuss proton-shuttle problem. The exponential scaling of the number of grid points and the associated number of electronic structure calculations needed for an accurate potential energy surface is also noted in Table 1. Furthermore, as is clear from eq 10, eq 9 is divided into multiple parts. While the number of $E_{level,0}$ calculations scales exponentially with system size, with one value needed for every grid point shown in Table 1, the number of fragment calculations scales within a reduced dimensional space. Our goal here is to reduce the steep algebraic scaling of each electronic structure calculation and to reduce the number of such electronic structure calculations needed to obtain the potential surface, through the sampling measures discussed in Section S2. Furthermore, the family of electronic structure calculations captured using eq 10, to be performed on the set of sampled points, are efficiently parallelized using a python-based job-scheduler library,[106] making it feasible to obtain post-Hartree−Fock level surfaces for the high-dimensional cases presented here.
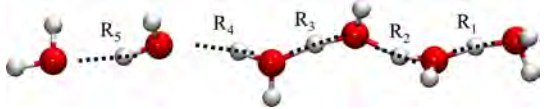
For the two-dimensional grid, referred to as $(R_1, R_2)$ in Table 1, a range of sample points is considered. These include sampling data ranging from 1 to 20% of the grid points, and these sample points are chosen from a uniformly distributed two-dimensional grid, defined using 99 grid points along each of the quantum nuclear dimensions shown on the figure in Table 1. This amounts to nearly 10,000 grid points, with a grid spacing of approximately 0.008 Å where good quality estimates of the electronic structure energies are needed. The question addressed in Table 2 is whether the sampling technique will allow us to perform this task with as little as 100 calculations (i.e., 1% of the total number of grid points), delivering, in the process, an acceptable accuracy in the sub-kcal/mol range for the potential surface, where the error is gauged using the estimates presented in Appendix D. A range of sampling points is considered for the two-dimensional case to gauge the computational advantage afforded by the sampling protocol, and clearly, as seen in Table 2, the errors are in the sub-kcal/mol range. Table 2 provides (a) the quantum-nuclear ground-state ($\psi_0$) weighted error on the full potential surface $\Delta\epsilon_{V,\omega}$ from eq D1 in Appendix D, with additional details in Section S5, (b) the Boltzmann factor-weighted error on the potential surface, $\Delta\epsilon_{V,\omega}$ from eq D1 in Appendix D, and (c) the absolute errors in the multidimensional quantum-nuclear eigenenergies ($\{E_i\}$), from eq D2 for the two-dimensional, $(R_1, R_2)$, case. Clearly, many of these error estimates need the multidimensional quantum-nuclear eigenstates to be computed, and in all cases considered here, the complexity of obtaining the solution of such eigenvalue problems is compounded by its size, as reflected by Table 1. Hence, we use the Arnoldi iterative eigen-spectral approach.[190−192] The Arnoldi method

is a generalization of the Lanczos iterative diagonalization procedure, where repeated action of a function of an operator on some initial random vector yields a Krylov basis. It has been shown[190] that representing the original operator using the Krylov vector basis leads to an eigenvalue problem that rapidly converges to the exterior eigenvalues. Details regarding construction of the multidimensional Hamiltonian and efficient action of the same on large-sized vectors are described in Appendix E and ref 164.

All energetic measures of errors on the post-Hartree−Fock potential surfaces are well within the sub-kcal/mol range, in Table 1, at a much affordable computational cost. For example, the ground-state weighted errors, $\Delta\epsilon_{V,\omega}$, in Table 2 are in the range 0.11−0.23 kcal/mol where the maximum error is seen when the fewest number of sampling points is used. Similarly, the smallest sample set leads to the largest error in the Boltzmann-weighted errors, eigenenergy errors, and the eigenstate density errors. The ground-state weighted errors on the potential surface are less than 0.2 kcal/mol in all except for the smallest sample-size case, and hence, it is an acceptable error.

The errors in eigenenergies increase for higher eigenstates for each sample set. For example, the absolute errors in zero-point energy, first-excited-state eigenenergy, and second-excited-state eigenenergy for a grid sampling compression factor of 0.02 (i.e., sample points being 2% of the grid) are 0.03, 0.11, and 0.20 kcal/mol, respectively. These errors are small and acceptable, and it must also be noted that the thermal population of the second excited vibrational state is negligible.

In Table 3, we present the errors for the interpolated potential surface, $E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\mathbf{R})$, for the coupled three-, four-, and five-dimensional quantum mechanical treatment of the hydrogen nuclear problem. As noted earlier, the coupled, quantum treatment of nuclear degrees of freedom is a challenging problem in computational chemistry, due potentially to (a) the exponential scaling in the number of electronic structure calculations that needs to be performed with the increase in nuclear dimensions (see Table 1, total number of grid points), (b) the exponential scaling of the time evolution of a wave packet state describing such a coupled system (see Appendix E), and (c) the steep polynomial scaling of the electronic structure for each individual geometry (or grid points in Table 1). The tests in Table 3 provide direct evidence of the use of eqs 7−9 in providing accurate higher-dimensional, correlated potential energy surfaces, with scaling costs as described in Appendix C. For the two-, three-, and the four-dimensional results, the reference potential energy surfaces are determined at the MP2/6-31++G(d,p) level of theory on the full uniform grid, defined as in Table 1, without any sampling procedure and are compared with the results of eq 9. For the five-dimensional result, both the reference and the potentials computed using eq 9 are computed using the sampling approach using only 1% of the uniform grid defined in Table 1. Clearly, as seen from Table 3, all error estimates in obtaining MP2 level potential surfaces are well within the sub-kcal/mol accuracy range. In future publications, we will further evaluate the use of this method in performing accurate quantum nuclear dynamics studies in higher nuclear dimensions. Next, we provide an analysis of the sampling procedure through the radial distribution of sample points

**Table 3. Error Estimates for the 3D, 4D, and 5D Cases**[a]



|  | 2D[b] | 3D[b] | 4D[b] | 5D[b] |
|---|---|---|---|---|
|  | $(R_1, R_2)$[c] | $(R_1, R_2, R_3)$[c] | $(R_1, R_2, R_3, R_4)$[c] | $(R_1, R_2, R_3, R_4, R_5)$[c] |
|  | 0.01[d] | 0.03[d] | 0.01[d] | 0.01[d] |
| no. of diabatic states included | 6 | 18 | 18 | 18 |
| $\Delta\epsilon_{V,\omega}$[e] | 0.23 | 0.30 | 0.21 | 0.37 |
| $\Delta\epsilon_{V,\omega}$[f] | 0.15 | 0.15 | 0.12 | 0.20 |
| $\Delta\epsilon_{E\alpha}$[g] | 0.22 | 0.16 | 0.10 | 0.05 |

[a]The number of sample points dictates the number of electronic structure energies needed and hence depicts the complexity. Also, see Table 1. [b]Dimensionality of the quantum nuclear grid which represents, in this case, the number of coupled quantum nuclear (proton) dimensions. See the figure above. [c]2D: total number of grid points = 9801 and total number of sample points = 100. 3D: total number of grid points = 117,649 and total number of sample points = 3456. 4D: total number of grid points = 5,764,801 and total number of sample points = 41,472. 5D: total number of grid points = 6,436,343 and total number of sample points = 65,536. [d]Ratio of the number of sampling points (last column in Table 1) to the total number of grid points (penultimate column in Table 1). The numbers along the first row depict the extent of grid compression due to sampling. [e]Potential surface error estimates (in kcal/mol) as per eq D1. The quantity $\omega \equiv \rho_0$ and the ground-state wave packet density. [f]Potential surface error estimates (in kcal/mol) as per eq D1. The quantity $\omega$ is chosen to be the Boltzmann factor with a temperature of 300 K. [g]Eigenvalue error estimates (in kcal/mol) as per eq D2. Ground-state eigenvalue error.

explained in Section 4.1.1 below. We analyze the distribution of sample points for two-, three-, and the four-dimensional problems only because we computed a full-dimensional potential energy surface as a reference for these cases. In contrast, for the five-dimensional case, we used the sampling approach to calculate the reference potential; thus, we are excluding that case in the next section.

*4.1.1. Distribution of Sample Points in Higher Dimensions.* To further gauge the accuracy and efficiency of the

sampling approach, in conjunction with the graph-theoretic method, we next consider the distribution of sample points obtained from this technique. As noted above, the errors in potential are in the sub-kcal/mol range for chemical accuracy. While the error estimates discussed above do allow us to gauge the effectiveness of the sampling measures, it is also important to analyze the location of the chosen sampled grid points to probe the effectiveness of the sampling measures. For this purpose, we consider all cases ranging from the two-dimensional coupled quantum-nuclear problem, where the two nuclear dimensions labeled $(R_1, R_2)$ in Table 1 are considered simultaneously, to the four-dimensional coupled problem labeled $(R_1, R_2, R_3, R_4)$ in Table 1. It is well known that coupled behavior of quantum nuclear dimensions, especially for protons, is critical for a range of spectroscopic[193,194] and dynamical problems,[127] but the computational complexity grows exponentially with the number of nuclear degrees of freedom, making these studies prohibitive, especially when state-of-the-art post-Hartree−Fock level surfaces are also desired. In studying such complex problems, it is crucial to sample the lower energy regions of the potential well enough to capture critical quantum nuclear effects. Thus, along with the errors on potential surfaces (using estimates in Appendix D), the errors in quantum nuclear effects such as eigenenergies and quantum nuclear eigenfunctions (see Appendix E) are a direct result of the choice and location of sampling performed here. For example, the errors in *low*-energy regions contribute more to the ground-state weighted error computed using eq D1.

Figure 4 provides an analysis of the spatial distribution of sampled grid points for all cases, where the plots within the different subfigures correspond to the different dimensions, such as $(R_1, R_2)$ for 2D, $(R_1, R_2, R_3)$ for 3D, and so on. To allow a visual analysis of the distribution of grid points in higher-dimensional problems, we present radially averaged distributions of the sample points (blue bars), associated distribution of all grid points on the uniform regularly spaced grid (red bars) used for interpolation of the potentials (see Section S4), and the radially averaged distribution functions computed with respect to the distance from the minimum energy configuration represented as $\mathbf{R}_{min}$, according to
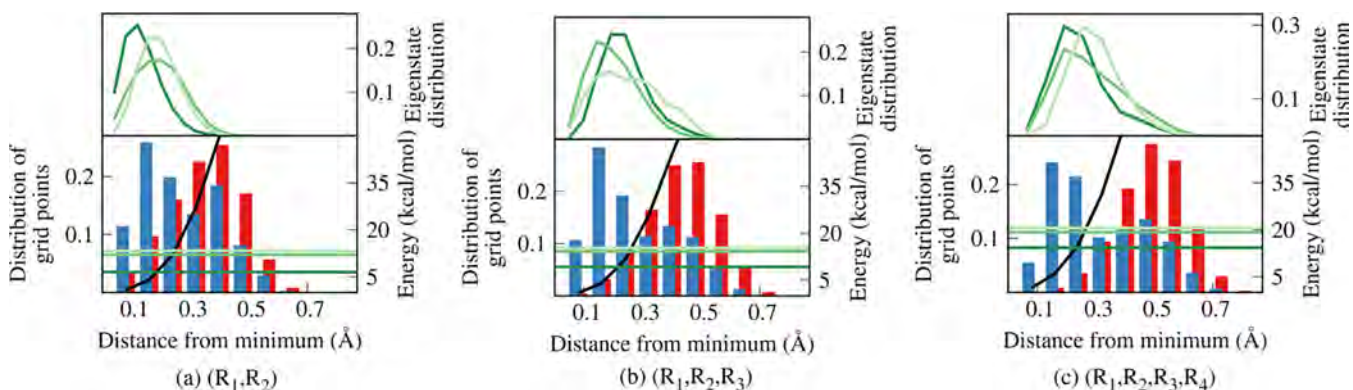


**Figure 4.** For each of the three subfigures presented above, the top panel shows the radially averaged eigenstate density distribution for the first three eigenstate probability densities, as obtained from eq 13. The bottom panel in each subfigure shows, in blue, the radial distribution of the fraction of sample points, again computed using eq 13, the radial distribution of the fraction of points on a uniform grid (red bars, left vertical axis), and the radially averaged potential energy, eq 14 (black line, right vertical axis). The first three eigenenergies are also shown using the horizontal line using the same colors as those for the eigenstates in the top panel.

$$\langle f(\|\mathbf{R} - \mathbf{R}_{\min}\|)\rangle_{\mathbf{R}} = \int d\Theta \, f(\|\mathbf{R} - \mathbf{R}_{\min}\|; \Theta) \tag{13}$$

where we have represented the multidimensional space as $\{\|\mathbf{R} - \mathbf{R}_{\min}\|; \Theta\}$, with $\|\mathbf{R} - \mathbf{R}_{\min}\|$ depicting the radial distance from the minimum potential energy region and $\Theta$ providing the appropriate cumulative, multidimensional orthogonal set of variables. For example, the radially averaged potential energy surface $\langle E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\mathbf{R})\rangle$ (black trace in the subfigures of Figure 4)

$$\langle E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\mathbf{R})\rangle = \int d\Theta \, E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\|\mathbf{R} - \mathbf{R}_{\min}\|; \Theta) \tag{14}$$

as a function of distance from the lowest energy configuration in the bottom panels of the three subfigures is shown in Figure 4. The three lowest eigenvalues and eigenfunctions are also marked with horizontal green lines of different shades in each case. We note that the blue bars showing the radial distribution of the fraction of sample points are skewed toward the lower energy regions as compared to the full grid represented using the red bars. Furthermore, the blue bars appear to provide adequate sampling in regions where the eigenstates have larger populations, as may be seen by visually correlating the radial population of eigenstates provided immediately above in the three subfigures in Figure 4, also computed as per eq 13.

## 5. CONCLUSIONS

Over a recent set of publications, we have shown how graph-theoretic methods combined with ONIOM and many-body theory can be used to perform (a) accurate post-Hartree–Fock AIMD at the DFT cost for medium- to large-sized molecular clusters, (b) hybrid DFT electronic structure calculations for condensed-phase simulations at the cost of pure density functionals, (c) reduced cost on-the-fly basis extrapolation for gas-phase AIMD and condensed-phase studies, and (d) accurate post-Hartree–Fock level potential energy surfaces at the DFT cost for quantum nuclear effects. The salient features of our approach are ONIOM-like in that (i) the full-system (cluster or condensed phase) calculation is performed at a lower level of theory (pure DFT for the condensed phase or hybrid DFT for molecular systems) and (ii) this approximation is improved through a correction term that captures all many-body interactions up to any given order within a higher level of theory (hybrid DFT for the condensed phase and CCSD or MP2 for cluster), combined through graph-theoretic methods. Specifically, a region of chemical interest is coarse-grained into a set of nodes, and these nodes are then connected to form edges based on a given definition of local envelope (or threshold) of interactions. The nodes and edges together define a graph, which forms the basis for developing the many-body expansion discussed here, embedded within an ONIOM scheme, but using higher-order simplexes that are composed from the graphical description. In general, the effect of adding a higher-order correction on the accuracy of potentials diminishes after the consideration of a few low-rank simplexes, and hence, the required fragments are far smaller than the full molecular system. In fact, the rapid convergence of the formalism here, as a function of many-body rank, has been previously demonstrated for condensed systems[115] and molecular clusters.[106]

However, when nuclear quantum effects are critical, multiple graphical descriptions are simultaneously needed to describe the potential energy surface. In this paper, we provide a general variational approach that combines multiple graphical descriptions of molecular systems to obtain accurate post-Hartree–Fock level potential surfaces at the DFT cost. The approach also potentially allows a drastic reduction in the computational complexity involved in computing accurate multidimensional potential surfaces. Due to the demonstrated rapid convergence of the approach with increasing rank,[106,115] the fragments considered for many-body expansions contain far fewer nuclear degrees of freedom than that of the full molecular system and hence span a much smaller portion of the conformational subspace. This provides a substantial reduction in the exponential scaling, and it is shown that the number of electronic structure calculations needed in $\mathcal{D}$ quantum dimensions reduces from $O(\mathcal{N}^{\mathcal{D}})$ to $O(\mathcal{N}^{(\mathcal{R}+1)\mathcal{D}/\mathcal{F}})$ when the graph-theoretic expressions are truncated at order $\mathcal{R}$ and the molecular system is partitioned into $\mathcal{F}$ nodes before creating a graph from these nodes. In addition, we also reduce the number of potential energy calculations by orders of magnitude, through a multidimensional grid clustering/tessellation approach, which further reduces the computational complexity.

The method is demonstrated for the protonated water-wire systems, where coupled multidimensional quantum nuclear effects are thought to have a significant[189] impact on proton transfer dynamics. We utilize our approach to compute accurate multidimensional post-Hartree–Fock potential energy surfaces. In all cases, accurate post-Hartree–Fock potential surfaces, in agreement with regular post-Hartree–Fock methods to within sub-kcal/mol, are obtained in multiple dimensions, at a much-reduced computation cost.

## ■ APPENDIX A

**Explicit Forms of Eqs 2, 3, and 9 for Two-Body and Three-Body Interactions**

*Illustration of Connections to Two-Body Theory.* The choice of $\mathcal{R} = 1$ in eq 2 leads to

$$
\begin{aligned}
E_{\text{MBE},\mathcal{R}=1}^{\text{level,I}}(\mathbf{R}) &= (-1)^0 \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,I}}(\mathbf{R}) \left[ \sum_{m=0}^{1} (-1)^m p_\alpha^{0,m} \right] \\
&\quad + (-1)^1 \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,I}}(\mathbf{R}) [(-1)^1 p_\alpha^{1,1}] \\
&= \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,I}}(\mathbf{R}) [(-1)^0 p_\alpha^{0,0} + (-1)^1 p_\alpha^{0,1}] \\
&\quad - \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,I}}(\mathbf{R}) [-p_\alpha^{1,1}] \\
&= \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,I}}(\mathbf{R}) [p_\alpha^{0,0} - p_\alpha^{0,1}] - \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,I}}(\mathbf{R}) [-p_\alpha^{1,1}]
\end{aligned}
\tag{A1}
$$

where the quantities $p_\alpha^{0,0}$ and $p_\alpha^{0,1}$ are the number of times the $\alpha$th node appears in the set of all nodes ($\mathbf{V}_0$) and all edges ($\mathbf{V}_1$), respectively; similarly, $p_\alpha^{1,1}$ is the number of times the $\alpha$th edge appears inside the set of all edges. Since the sets such as $\mathbf{V}_0$, $\mathbf{V}_1$, and so forth do not contain any duplicates, $p_\alpha^{0,0} = p_\alpha^{1,1} = 1$. Additionally, the first term in eq A1

$$\sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,I}}(\mathbf{R}) [p_\alpha^{0,0}] \equiv \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,I}}(\mathbf{R})$$

provides the one-body contribution to the approximation to the potential energy surface using many-body expansions. The remaining portion of eq A1

$$\sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,I}}(\mathbf{R})[-p_\alpha^{0,1}] - \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,I}}(\mathbf{R})[-p_\alpha^{1,1}]$$

provides the two-body correction, and thus

$$E_{\text{MBE},\mathcal{R}=1}^{\text{level,I}}(\mathbf{R}) = \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,I}}(\mathbf{R}) + \left\{ \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,I}}(\mathbf{R}) \right.$$
$$\left. - \sum_{\alpha \in \mathbf{V}_0} p_\alpha^{0,1} E_{\alpha,0}^{\text{level,I}}(\mathbf{R}) \right\} \tag{A2}$$

The many-body expansion described in eq A1 is combined with the ONIOM scheme computed at two levels of theory labeled "level, 1" and "level, 0" in eq 3. Considering $\mathcal{R} = 1$ in eq 3 similarly leads to

$$E_{\mathcal{R}=1}^{1,0}(\mathbf{R}) = E^{\text{level,0}}(\mathbf{R}) + \left( \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,1}}(\mathbf{R})[p_\alpha^{0,0} - p_\alpha^{0,1}] \right.$$
$$\left. - \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,1}}(\mathbf{R})[-p_\alpha^{1,1}] \right)$$
$$- \left( \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,0}}(\mathbf{R})[p_\alpha^{0,0} - p_\alpha^{0,1}] - \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,0}}(\mathbf{R})[-p_\alpha^{1,1}] \right)$$
$$= E^{\text{level,0}}(\mathbf{R}) + \sum_{\alpha \in \mathbf{V}_0} \Delta E_{\alpha,0}^{1,0}(\mathbf{R})[p_\alpha^{0,0} - p_\alpha^{0,1}]$$
$$- \sum_{\alpha \in \mathbf{V}_1} \Delta E_{\alpha,1}^{1,0}(\mathbf{R})[-p_\alpha^{1,1}]$$
$$= E^{\text{level,0}}(\mathbf{R}) + \sum_{\alpha \in \mathbf{V}_0} \Delta E_{\alpha,0}^{1,0}(\mathbf{R}) + \left\{ \sum_{\alpha \in \mathbf{V}_1} \Delta E_{\alpha,1}^{1,0}(\mathbf{R}) \right.$$
$$\left. - \sum_{\alpha \in \mathbf{V}_0} p_\alpha^{0,1} \Delta E_{\alpha,0}^{1,0}(\mathbf{R}) \right\} \tag{A3}$$

Thus, a many-body expansion is constructed on the difference in level, 1 and level, 0 energies as recommended from ONIOM. Correspondingly, eq 9 takes the form

$$E_{\mathcal{R}=1,\{\mathcal{G}_\beta\}}(\mathbf{R}) = E^{\text{level,0}}(\mathbf{R}) + \sum_\beta v_\beta(\mathbf{R}) \left( \sum_{\alpha \in \mathbf{V}_0^\beta} \Delta E_{\alpha,0}^{1,0}(\mathbf{R}) \right.$$
$$\left. + \left\{ \sum_{\alpha \in \mathbf{V}_1^\beta} \Delta E_{\alpha,1}^{1,0}(\mathbf{R}) - \sum_{\alpha \in \mathbf{V}_0^\beta} p_\alpha^{0,1} \Delta E_{\alpha,0}^{1,0}(\mathbf{R}) \right\} \right) \tag{A4}$$

where the fact that each graph may have a different set of nodes and edges is depicted using the notation $\mathcal{G}_\beta \equiv \{\mathbf{V}_0^\beta; \mathbf{V}_1^\beta\}$.

**Illustration of Connections to Three-Body Theory**

The choice of $\mathcal{R} = 2$ in eq 2 leads to

$$E_{\text{MBE},\mathcal{R}=2}^{\text{level,I}}(\mathbf{R}) = (-1)^0 \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,I}}(\mathbf{R}) \left[ \sum_{m=0}^{2} (-1)^m p_\alpha^{0,m} \right]$$
$$+ (-1)^1 \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,I}}(\mathbf{R}) \left[ \sum_{m=1}^{2} (-1)^m p_\alpha^{1,m} \right]$$
$$+ (-1)^2 \sum_{\alpha \in \mathbf{V}_2} E_{\alpha,2}^{\text{level,I}}(\mathbf{R})[(-1)^2 p_\alpha^{2,2}]$$
$$= \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,I}}(\mathbf{R})[(-1)^0 p_\alpha^{0,0} + (-1)^1 p_\alpha^{0,1} + (-1)^2 p_\alpha^{0,2}]$$
$$- \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,I}}(\mathbf{R})[(-1)^1 p_\alpha^{1,1} + (-1)^2 p_\alpha^{1,2}]$$
$$+ \sum_{\alpha \in \mathbf{V}_2} E_{\alpha,2}^{\text{level,I}}(\mathbf{R})[(-1)^2 p_\alpha^{2,2}]$$

and rearranging terms leads to

$$E_{\text{MBE},\mathcal{R}=2}^{\text{level,I}}(\mathbf{R}) = \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,I}}(\mathbf{R})[p_\alpha^{0,0} - p_\alpha^{0,1} + p_\alpha^{0,2}]$$
$$- \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,I}}(\mathbf{R})[-p_\alpha^{1,1} + p_\alpha^{1,2}] + \sum_{\alpha \in \mathbf{V}_2} E_{\alpha,2}^{\text{level,I}}(\mathbf{R})[p_\alpha^{2,2}]$$
$$= \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,I}}(\mathbf{R}) + \left\{ \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,I}} - \sum_{\alpha \in \mathbf{V}_0} p_\alpha^{0,1} E_{\alpha,0}^{\text{level,I}}(\mathbf{R}) \right\}$$
$$+ \left\{ \sum_{\alpha \in \mathbf{V}_2} E_{\alpha,2}^{\text{level,I}}(\mathbf{R}) - \sum_{\alpha \in \mathbf{V}_1} p_\alpha^{1,2} E_{\alpha,1}^{\text{level,I}}(\mathbf{R}) \right.$$
$$\left. + \sum_{\alpha \in \mathbf{V}_0} p_\alpha^{0,2} E_{\alpha,0}^{\text{level,I}} \right\} \tag{A5}$$

where the quantities $p_\alpha^{0,0}$, $p_\alpha^{0,1}$, and $p_\alpha^{1,1}$ hold the same meaning as in eq A1 defined above. Similarly, the quantities $p_\alpha^{0,2}$, $p_{\alpha'}^{1,2}$, and $p_{\alpha''}^{2,2}$ are the number of times the $\alpha$th node, the $\alpha'$th edge, and $\alpha''$th triangle appears in the set of all triangles, $\mathbf{V}_2$. As mentioned earlier, the sets of $\{\mathbf{V}_r\}$ do not contain any duplicates, and hence, $p_\alpha^{2,2} = 1$. The final bracketed term, $\{\cdots\}$, in eq A5

$$\sum_{\alpha \in \mathbf{V}_2} E_{\alpha,2}^{\text{level,I}}(\mathbf{R}) - \sum_{\alpha \in \mathbf{V}_1} p_\alpha^{1,2} E_{\alpha,1}^{\text{level,I}}(\mathbf{R}) + \sum_{\alpha \in \mathbf{V}_0} p_\alpha^{0,2} E_{\alpha,0}^{\text{level,I}}$$

is the three-body correction to the two-body many-body expansion, in eq A2. The many-body nature of eq 2 is very clear from eqs A2 and A5.

The choice of the many-body expansion for $\mathcal{R} = 2$ obtained in the eq A5 when utilized in eq 3 leads to

$$E_{\mathcal{R}=2}^{1,0}(\mathbf{R}) = E^{\text{level,0}}(\mathbf{R})$$
$$+ \left( \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,1}}(\mathbf{R})[p_\alpha^{0,0} - p_\alpha^{0,1} + p_\alpha^{0,2}] - \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,1}}(\mathbf{R}) \right.$$
$$\left. [-p_\alpha^{1,1} + p_\alpha^{1,2}] + \sum_{\alpha \in \mathbf{V}_2} E_{\alpha,2}^{\text{level,1}}(\mathbf{R})[p_\alpha^{2,2}] \right)$$
$$- \left( \sum_{\alpha \in \mathbf{V}_0} E_{\alpha,0}^{\text{level,0}}(\mathbf{R})[p_\alpha^{0,0} - p_\alpha^{0,1} + p_\alpha^{0,2}] - \sum_{\alpha \in \mathbf{V}_1} E_{\alpha,1}^{\text{level,0}}(\mathbf{R}) \right.$$
$$\left. [-p_\alpha^{1,1} + p_\alpha^{1,2}] + \sum_{\alpha \in \mathbf{V}_2} E_{\alpha,2}^{\text{level,0}}(\mathbf{R})[p_\alpha^{2,2}] \right)$$
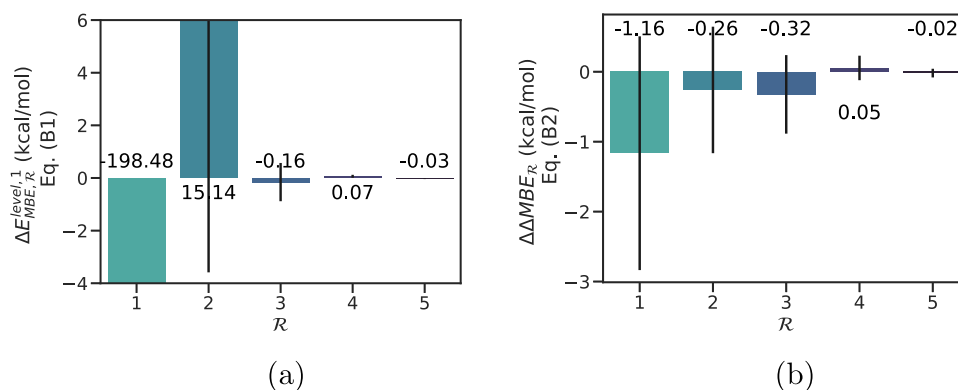
and rearranging terms leads to

(a)          (b)

**Figure 5.** Rapid convergence of eq 3 (b) as compared to eq 2 (a) in obtaining accurate CCSD energies. The data are obtained from all configurations from a 10 ps AIMD trajectory, for a protonated 21-water cluster computed in ref 106, which uses eq 3 as the form of energy. The horizontal axis $\mathcal{R}$ represents the rank of the graph and hence the order of MBE ($\mathcal{R} + 1$). For example, $\mathcal{R} = 2$ implies three-body interactions and so on. Hence, while eq 2 in the paper and consequently MBE-type approaches need $\mathcal{R} = 3$ (four-body terms) for sub-kcal/mol convergences, this is already the case where $\mathcal{R} = 2$ and lower provides much more satisfactory results when using eq 3.

$$E_{\mathcal{R}=2}^{1,0}(\mathbf{R}) = E^{\text{level},0}(\mathbf{R}) + \sum_{\alpha \in \mathbf{V}_0} \Delta E_{\alpha,0}^{1,0}(\mathbf{R})[p_\alpha^{0,0} - p_\alpha^{0,1} + p_\alpha^{0,2}]$$

$$- \sum_{\alpha \in \mathbf{V}_1} \Delta E_{\alpha,1}^{1,0}(\mathbf{R})[-p_\alpha^{1,1} + p_\alpha^{1,2}] + \sum_{\alpha \in \mathbf{V}_2} \Delta E_{\alpha,2}^{1,0}(\mathbf{R})[p_\alpha^{2,2}]$$

$$= E^{\text{level},0}(\mathbf{R}) + \sum_{\alpha \in \mathbf{V}_0} \Delta E_{\alpha,0}^{1,0}(\mathbf{R}) + \left\{ \sum_{\alpha \in \mathbf{V}_1} \Delta E_{\alpha,1}^{1,0}(\mathbf{R}) \right.$$

$$\left. - \sum_{\alpha \in \mathbf{V}_0} p_\alpha^{0,1} \Delta E_{\alpha,0}^{1,0}(\mathbf{R}) \right\}$$

$$+ \left\{ \sum_{\alpha \in \mathbf{V}_2} \Delta E_{\alpha,2}^{1,0}(\mathbf{R}) - \sum_{\alpha \in V_1} p_\alpha^{1,2} \Delta E_{\alpha,1}^{1,0} + \sum_{\alpha \in V_0} p_\alpha^{0,2} \Delta E_{\alpha,0}^{1,0}(\mathbf{R}) \right\}$$

(A6)

The ONIOM scheme defined using eq A6 with an MBE truncated at $\mathcal{R} = 2$ is calculated using a series of molecular graphs $\mathcal{G}_\beta$ considered as diabatic states, and an adiabatic state is obtained using eq 9 as

$$E_{\mathcal{R}=2,\{\mathcal{G}_\beta\}}(\mathbf{R}) = E^{\text{level},0}(\mathbf{R})$$

$$+ \sum_\beta v_\beta(\mathbf{R}) \left( \sum_{\alpha \in \mathbf{V}_0^\beta} \Delta E_{\alpha,0}^{1,0}(\mathbf{R})[p_\alpha^{0,0} - p_\alpha^{0,1} + p_\alpha^{0,2}] \right.$$

$$\left. - \sum_{\alpha \in \mathbf{V}_1^\beta} \Delta E_{\alpha,1}^{1,0}(\mathbf{R})[-p_\alpha^{1,1} + p_\alpha^{1,2}] + \sum_{\alpha \in \mathbf{V}_2^\beta} \Delta E_{\alpha,2}^{1,0}(\mathbf{R})[p_\alpha^{2,2}] \right)$$

$$= E^{\text{level},0}(\mathbf{R}) + \sum_\beta v_\beta(\mathbf{R}) \left( \sum_{\alpha \in \mathbf{V}_0^\beta} \Delta E_{\alpha,0}^{1,0}(\mathbf{R}) \right.$$

$$+ \left\{ \sum_{\alpha \in \mathbf{V}_1^\beta} \Delta E_{\alpha,1}^{1,0}(\mathbf{R}) - \sum_{\alpha \in \mathbf{V}_0^\beta} p_\alpha^{0,1} \Delta E_{\alpha,0}^{1,0}(\mathbf{R}) \right\}$$

$$\left. + \left\{ \sum_{\alpha \in \mathbf{V}_2^\beta} \Delta E_{\alpha,2}^{1,0}(\mathbf{R}) - \sum_{\alpha \in \mathbf{V}_1^\beta} p_\alpha^{1,2} \Delta E_{\alpha,1}^{1,0} + \sum_{\alpha \in \mathbf{V}_0^\beta} p_\alpha^{0,2} \Delta E_{\alpha,0}^{1,0}(\mathbf{R}) \right\} \right)$$

(A7)

where, again, the embedded many-body nature of the approach is clear.

### ■ APPENDIX B

**Comparison of Convergence Properties of Eqs 2 and 3 as a Function of $\mathcal{R}$**

Our previous studies in refs 115 and 106 have shown that for a variety of problems and for a variety of nuclear configurations, obtained from AIMD calculations that use eq 3, the resultant energy and dynamical correlation functions converge very quickly to the correct, correlated electronic structure result obtained from CCSD and MP2 levels of theory on the full system. Thus, all our previous results in refs 99, 102−104, 106, 107, and 113−115 have always been compared with the full-system results computed at the target post-Hartree−Fock correlated level of theory for dynamics and for potential surfaces. This, of course, includes all long-range electronic effects. The convergence of eq 3 which is the ONIOM-style expression with embedded many-body expansions is drastically improved as compared to standard many-body theory. Figures S14 through S16 and S20 in ref 115 and Figures 5 and 6 in ref 106 highlight this idea. We provide here Figures 5 and 6 to summarize this convergent behavior. These figures are obtained from 10 ps-long AIMD trajectories obtained in ref 106 for the protonated 21-water cluster, $(H_2O)_{21}H^+$.

As highlighted in Appendix A, eq 2 is basically a many-body expansion that is made efficient due to the use of graph-theory implementations within Python. However direct use of eq 2 and hence many-body theory does not converge rapidly to the post-Hartree−Fock result, as compared to eq 3. This is seen from Figure 5. Here

$$\Delta E_{\text{MBE},\mathcal{R}}^{\text{level},\text{I}} \equiv E_{\text{MBE},\mathcal{R}}^{\text{level},\text{I}} - E_{\text{MBE},\mathcal{R}-1}^{\text{level},\text{I}}$$

(B1)

where the two terms on the right are two successive approximations from eq 2 and in Figure 5b, and here

$$\Delta\Delta\text{MBE}_{\mathcal{R}} \equiv \Delta\text{MBE}_{\mathcal{R}}^{\text{level},1;\text{level},0} - \Delta\text{MBE}_{\mathcal{R}-1}^{\text{level},1;\text{level},0}$$

(B2)

and $\Delta\text{MBE}_{\mathcal{R}}^{\text{level},1;\text{level},0}$ is the correction term in eq 3. Certainly, eq 3 converges faster as compared to eq 2. This result not only holds for clusters of different sizes but also for periodic condensed phase systems, as discussed in ref 115. In Figure 6, we also show that eq 3 converges to the best possible result, in this case obtained using six-body interactions obtained with $\mathcal{R} = 5$ in a systematic manner.
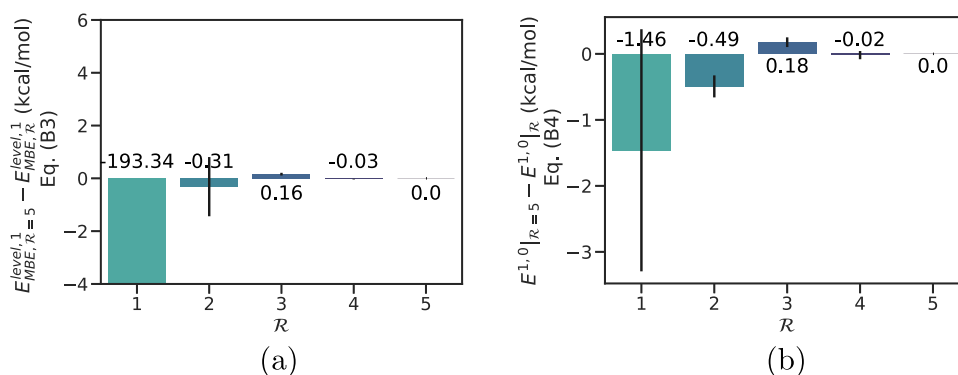
**Figure 6.** Similar to Figure 5, but now, the accuracy of each expression is considered with respect to the best case $\mathcal{R} = 5$ result for each case. Clearly, the convergence of the ONIOM result is superior at lower levels of $\mathcal{R}$ as compared to the MBE result.

The vertical axis in Figure 6a presents the difference between eq 2 at specific values of $\mathcal{R}$ with respect to the maximum computed value $\mathcal{R} = 5$ which includes all many-body interactions until six-body terms. Thus

$$E_{\text{MBE},\mathcal{R}=5}^{\text{level},1} - E_{\text{MBE},\mathcal{R}}^{\text{level},1} \tag{B3}$$

Similarly, Figure 6b contains the analogous quantity for eq 3

$$E_{\mathcal{R}=5}^{1,0} - E_{\mathcal{R}}^{1,0} \tag{B4}$$

## ■ APPENDIX C

### Intrinsic Computational Scaling of Eq 9 and the Effect of Sampling

The number of electronic structure calculations needed to compute the potential surface, $E_{\{\mathcal{G}_\beta\}}(\mathbf{R})$, obtained from the associated set of fragment energies, for a set of nuclear geometries within $\mathbf{R}$, scales exponentially with nuclear dimensions. For example, let us consider $\mathcal{D}$-nuclear dimensions with each dimension containing $\mathcal{N}$ discretizations where potential energy surface values are to be evaluated. Thus, in total, $\mathcal{N}^{\mathcal{D}}$ potential energy calculations are to be performed. Furthermore, the computational scaling for each electronic structure calculation is steeply algebraic when accurate post-Hartree–Fock results are desired. Through refs 99, 102–104, 106, 107, and 113–115 and the vast resources in the fragmentation literature,[34,35,37,39–41,43,50,52,53,56,58,59,61–63,65–67,69,71,72,74–76,78,81,83,85,98,116–120,195–199] a plausible solution to the steep algebraic scaling problem has been developed.

If we directly use eq 9, one may imagine a total of $\mathcal{F}$ coarse-grained nodes being created from the graphical decomposition, where, for the sake of simplicity, we assume that each node contains $\mathcal{N}_e$ electronic basis functions. Furthermore, if we assume that the $\mathcal{D}$-nuclear dimensions are equally partitioned into the $\mathcal{F}$ nodes (see Figure 2), the number of potential energy surface values to be evaluated for each node is of the order of $\mathcal{N}^{\mathcal{D}/\mathcal{F}} \ll \mathcal{N}^{\mathcal{D}}$. Thus, the overall complexity for direct evaluation of eq 9 becomes

$$\{\mathcal{N}^{\mathcal{D}} \cdot O([\mathcal{F}\mathcal{N}_e]^{\text{L}}) + \mathcal{N}^{(\mathcal{R}+1)\mathcal{D}/\mathcal{F}} \cdot [O([(\mathcal{R}+1)\mathcal{N}_e]^{\text{L}}) + O([(\mathcal{R}+1)\mathcal{N}_e]^{\text{H}})]\} \tag{C1}$$

here, the superscripts H and L indicate the raw algebraic scaling of the level, 1 and level, 0 calculations, respectively.

Clearly, the computational complexity in eq C1 arises from two sources. However, the second term of eq C1, $\{\mathcal{N}^{(\mathcal{R}+1)\mathcal{D}/\mathcal{F}} \cdot [O([(\mathcal{R}+1)\mathcal{N}_e]^{\text{L}}) + O([(\mathcal{R}+1)\mathcal{N}_e]^{\text{H}})]\}$, is much smaller in complexity than the full evaluation in the absence of molecular fragmentation, that is, $\mathcal{N}^{\mathcal{D}} \cdot O([\mathcal{F}\mathcal{N}_e]^{\text{H}})$, as highlighted in Figure 2 using brown-colored simplexes. The complexity of the higher level electronic structure portion of the molecular potential surface evaluations is reduced through the use of eq 9 to the order of

$$\{\mathcal{N}^{\mathcal{D}} \cdot O([\mathcal{F}\mathcal{N}_e]^{\text{H}}) \rightarrow \mathcal{N}^{(\mathcal{R}+1)\mathcal{D}/\mathcal{F}} \cdot [O([(\mathcal{R}+1)\mathcal{N}_e]^{\text{H}})]\} \tag{C2}$$

In Section S2, we introduced a sampling measure that is used to reduce the number of calculations. The discrete sample obtained using the algorithm discussed in Section S2 is the set of nuclear configurations where electronic structure computations are to be performed to obtain the quantities in eq 10. As seen in Section 4, the size of the discrete sample set is very much smaller than $\mathcal{N}^{\mathcal{D}}$, for $\mathcal{D}$-nuclear dimensions with each dimension containing $\mathcal{N}$ discretizations where potential energy surface values are to be evaluated (see Tables 1–3). This further reduces the scaling of the algorithm, as discussed in eq C1, to

$$\mathcal{N}^{\mathcal{D}} \cdot O([\mathcal{F}\mathcal{N}_e]^{\text{L}}) + \mathcal{N}^{(\mathcal{R}+1)\mathcal{D}/\mathcal{F}} \cdot [O([(\mathcal{R}+1)\mathcal{N}_e]^{\text{L}})$$
$$+ O([(\mathcal{R}+1)\mathcal{N}_e]^{\text{H}})]$$
$$\xrightarrow{\text{sampling}} \{\mathcal{N}_s \cdot O([\mathcal{F}\mathcal{N}_e]^{\text{L}}) + \mathcal{N}_s^{\mathcal{R}+1/\mathcal{F}} \cdot$$
$$[O([(\mathcal{R}+1)\mathcal{N}_e]^{\text{L}}) + O([(\mathcal{R}+1)\mathcal{N}_e]^{\text{H}})]\} \tag{C3}$$

where the quantity $\mathcal{N}_s$ represents the size of the sample set obtained using the discussion in Section S2, the lower scaling, full system part of the calculation is made efficient by sampling $\mathcal{N}^{\mathcal{D}} \cdot O([\mathcal{F}\mathcal{N}_e]^{\text{L}}) \rightarrow \mathcal{N}_s \cdot O([\mathcal{F}\mathcal{N}_e]^{\text{L}})$, and the higher scaling portion indeed undergoes a large reduction in complexity through

$$O(\mathcal{N}^{\mathcal{D}}) \xrightarrow{\text{eq 9}} O(\mathcal{N}^{(\mathcal{R}+1)\mathcal{D}/\mathcal{F}}) \xrightarrow{\text{sampling}} O(\mathcal{N}_s^{\mathcal{R}+1/\mathcal{F}}) \tag{C4}$$

and $\mathcal{N}_s^{\mathcal{R}+1/\mathcal{F}} \ll \mathcal{N}^{(\mathcal{R}+1)\mathcal{D}/\mathcal{F}}$.

## ■ APPENDIX D

### Error Estimates Used to Compute the Accuracy of the Multitopology Potential Surfaces

We introduce a series of error estimates used to gauge the accuracy of the potentials computed by the approach introduced in Section 2. Errors for the interpolated potential are computed using the absolute deviation of $E_{\{\mathcal{G}_\beta\}}(\mathbf{R})$ from its reference level, full-system level, 1 calculation

$$\Delta\epsilon_{V,\omega} = \int d\mathbf{R}\, \omega(\mathbf{R})|E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\mathbf{R}) - E_{\text{ref}}(\mathbf{R})| \tag{D1}$$

where $E_{\text{ref}}(\mathbf{R})$ is the potential energy computed at level, 1 of electronic structure theory at configuration $\mathbf{R}$. The quantity $\omega(\mathbf{R})$ acts as a filter that has a higher weight in the lower-energy critical regions of the potential. We have used two different kinds of approximations for $\omega(\mathbf{R})$. (a) We have used the Boltzmann approximation to construct these weights with $\omega(\mathbf{R}) = \frac{1}{N_0}\exp(-E_{\text{ref}}(\mathbf{R})/kT)$, where $T$ is the room temperature (300 K) and $N_0 \equiv \int \omega(\mathbf{R})d\mathbf{R}$ is a normalization constant. (b) We have also used the quantum nuclear ground eigenstate density, associated with the potential surface as an approximation to $\omega(\mathbf{R})$, in computing such errors. Obtaining the quantum nuclear ground state is carried out using the Arnoldi iterative diagonalization[190−192] scheme presented in Appendix E. This scheme provides the eigenenergies corresponding to the potential surfaces computer here and is used here to gauge the errors in potential in physically important regions and complement the results from eq D1. The absolute error in eigenenergies is

$$\Delta\epsilon_{E_\alpha} = |E_\alpha^{\{\mathcal{G}_\beta\}} - E_\alpha^{\text{ref}}| \tag{D2}$$

where again $E_\alpha^{\{\mathcal{G}_\beta\}}$ and $E_\alpha^{\text{ref}}$ are $\alpha$-th eigenenergies as previously defined for the eigenstates.

## ■ APPENDIX E

### Computing Eigenstates for the High-Dimensional Case Using Arnoldi Decomposition

To obtain the eigenstates ($\{\psi_i\}$) and their eigenenergies ($\{E_i\}$) used in eqs D1 and D2 for computing the ground-state weighted potential energy surface errors and the errors in eigenenergies, we form the Hamiltonian matrix corresponding to the multidimensional quantum nuclear degrees of freedom in the coordinate representation using a uniform grid basis, $\{x_i\}$, as

$$\hat{H}_{x_i,x_i'} = \sum_{i=1}^{N} \hat{K}_{x_i,x_i'} + E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\overline{x}) \tag{E1}$$

where the potential energy $E_{\mathcal{R},\{\mathcal{G}_\beta\}}(\overline{x})$ is that introduced in eq 9 in the paper and $\overline{x} = (x_1, x_2, ..., x_i, ..., x_N)$. The kinetic energy operator along each dimension, $\hat{K}_{x_i,x_i'}$, is represented using distributed approximating functionals (DAFs)[200−202] as

$$\hat{K}_{x_i,x_i'} = \frac{1}{\sigma\sqrt{2\pi}}\left(\frac{-1}{\sqrt{2}\,\sigma}\right)^2 \exp\left(-\frac{(x_i - x_i')^2}{2\sigma^2}\right)$$
$$\sum_{n=0}^{M/2}\left(\frac{-1}{4}\right)^n \frac{1}{n!}H_{2n+2}\left(\frac{x_i - x_i'}{\sqrt{2}\,\sigma}\right) \tag{E2}$$

where $H_{2n+2}$ denotes the even-order Hermite polynomials. The DAF approach has been well reviewed in the literature.[200,201] Thus, we only present a brief summary of the salient features in DAF representation in Appendix E.

Given the form of the multidimensional quantum nuclear Hamiltonian in eq E1, the Arnoldi scheme requires repeated action of a function of the Hamiltonian $f(\hat{H})$ on some random (or preconditioned) initial state which yields a Krylov basis, $\{|\chi_i\rangle:|\chi_i\rangle = f(\hat{H}|\chi_{i-1}\rangle)\}$. It is known[190] that such a basis when used to represent the Hamiltonian yields a tridiagonal (for Hermitian matrices) or an upper Heisenberg (for non-Hermitian matrices), each of which quickly provides the exterior eigenstates of the original matrix. Due to the banded-Toeplitz nature of the kinetic energy operator in eq E2, the repeated action of the Hamiltonian can be performed in an extremely efficient manner, as discussed in ref 203.

*"Distributed Approximating Functional" (DAF) Representation for the Kinetic Energy Operator (Eq E2).* The starting point is to expand a quantum wave packet, which is a function of the nuclear degrees of freedom, at time $t = 0$ using a local set of symmetric fitting functions, $a(x − x_i)$, such that

$$\chi(x; t = 0) = \sum_i \Delta x_i a(x - x_i)\chi(x_i; t = 0) \tag{E3}$$

where $\Delta x_i$ is the grid spacing (not, in general, uniform). The functions $a(x − x_i)$ are *local* fitting functions, the choice for which may, in general, depend upon the point $x_i$. One of the most common directions at this point is to assume that $a(x − x_i) \equiv \delta(x − x_i) \equiv \langle x|x_i\rangle$ is a suitable approximation to the Dirac $\delta$ function. Subsequent resolution of the identity in terms of some complete set of basis functions leads to a representation of the wave packet in that basis. The DAF approximation differs from these approaches by assuming that a suitable *local* representation[200,204] can be directly constructed for $a(x − x_i)$, that is

$$a(x - x_i) \equiv a_N(x - x_i; \sigma)$$
$$= \sum_{n}^{N} b_n H_n\left(\frac{x - x_i}{\sqrt{2}\,\sigma}\right)\exp\left[-\frac{(x - x_i)^2}{2\sigma^2}\right] \tag{E4}$$

where $H_n(x)$ denotes the Hermite polynomials generated from Gaussian according to

$$H_n(y)\exp[-y] = (-1)^n \frac{d^n}{dy^n}\exp[-y] \tag{E5}$$

Note that eq E4 is different from the expression obtained using a standard basis set approximation for $a(x − x_i)$ wherein the appropriate expression would be

$$a(x - x_i) \equiv \delta(x - x_i)$$
$$= \sum_n c_n\{\exp[-x^2/2]H_n(x)\}\{\exp[-x_i^2/2]H_n(x_i)\} \tag{E6}$$

with $c_n = (\sqrt{\pi}2^n n!)^{-1}$. Note that eq E6 is separable in $x$ and $x_i$, whereas eq E4 only depends on $(x − x_i)$. The *local spectral*[205] form in eq E4 has many computational advantages not the least of which is the fact that eq E4 yields a banded-Toeplitz matrix at *any* level of approximation. The choice of Hermite functions here is by no means a requirement; it is however a convenient choice. Using the orthogonality of the

Hermite functions and the fact that $a(x - x_i)$ must be symmetric with respect to interchange of $x$ and $x_i$ (since it approximates the Dirac $\delta$ function), one obtains

$$b_{2n+1} = 0$$

$$b_{2n} = \frac{1}{\sigma\sqrt{2\pi}}\left(-\frac{1}{4}\right)^n \frac{1}{n!} \quad \text{(E7)}$$

where we have used the identity[206]

$$\int dy \, \exp[-y^2] H_n(y) H_m(y) \equiv \delta_{n,m} 2^m m! \sqrt{\pi} \quad \text{(E8)}$$

thus resulting in eq E2. The variables $M$ and $\sigma$ determine the accuracy and width (or computational efficiency), respectively, of the DAF. It has been shown[26,200,207] that these parameters are not independent, and for a given value of $M$, there exists a $\sigma$ that provides optimal accuracy for the propagation. The accuracy of this method in conjunction with ab initio dynamics has been benchmarked in ref 26. For an approximation controlled by the choice of parameters $M$ and $\sigma$, eq E2 only depends on the quantity $(x - x')$, that is, the distance between points in the coordinate representation, and goes to zero as this quantity becomes numerically large due to the Gaussian prefactor. This yields a banded matrix approximation to eq E2 for any $M$ and $\sigma$. Furthermore, on account of its dependence on $(x - x')$, a matrix representation of eq E2 has the property that all diagonal elements of this matrix are equal; similarly, all $n$-th super (and sub) diagonal elements are the same. Such a matrix is called a Toeplitz matrix. The dependence on $(x - x')$ also implies a translational symmetry reminiscent of wavelet theories.[208–211]

We finally note that the DAF approach differs from other approaches that use Hermite functions to represent the wave packet[212,213] based on Heller's Gaussian wave packet formalism.[214] Within these formalisms,[212,213] a locally harmonic approximation to the potential[214] allows the reduction of the time-dependent Schrödinger equation to classical-like equations to propagate the width and center of the Gaussian wave packets. In our case, no assumption is made on the nature of the potential.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00065.

Brief description of the grids used to compute the multidimensional potential energy surfaces, further details on the sampling measure, and details on the interpolation method used in this work (PDF)

## AUTHOR INFORMATION

### Corresponding Author

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

### Authors

**Anup Kumar** − *Department of Chemistry and Department of Physics, Indiana University, Bloomington, Indiana 47405, United States*

**Nicole DeGregorio** − *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.1c00065

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

## REFERENCES

(1) Iannuzzi, M.; Parrinello, M. Proton Transfer in Heterocycle Crystals. *Phys. Rev. Lett.* **2004**, *93*, 025901.

(2) Tse, Y.-L. S.; Herring, A. M.; Kim, K.; Voth, G. A. Molecular dynamics simulations of proton transport in 3M and nafion perfluorosulfonic acid membranes. *J. Phys. Chem. C* **2013**, *117*, 8079−8091.

(3) Lin, I.-H.; Lu, Y.-H.; Chen, H.-T. Nitrogen-doped $C_{60}$ as a robust catalyst for CO oxidation. *J. Comput. Chem.* **2017**, *38*, 2041−2046.

(4) Pomès, R.; Roux, B. Structure and dynamics of a proton wire: a theoretical study of H+ translocation along the single-file water chain in the gramicidin A channel. *Biophys. J.* **1996**, *71*, 19−39.

(5) Pomès, R.; Roux, B. Theoretical Study of $H^+$ Translocation Along a Model Proton Wire. *J. Phys. Chem.* **1996**, *100*, 2519.

(6) Decornez, H.; Drukker, K.; Hammes-Schiffer, S. Solvation and Hydrogen-Bonding Effects on Proton Wires. *J. Phys. Chem. A* **1999**, *103*, 2891.

(7) Brewer, M. L.; Schmitt, U. W.; Voth, G. A. The formation and dynamics of proton wires in channel environments. *Biophys. J.* **2001**, *80*, 1691−1702.

(8) Teeter, M. M. Water Structure of a Hydrophobic Protein at Atomic Resolution: Pentagon Rings of Water Molecules in Crystals of Crambin. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 6014.

(9) Neidle, S.; Berman, H. M.; Shieh, H. S. Highly Structured Water Network in Crystals of a Deoxydinucleoside-Drug Complex. *Nature* **1980**, *288*, 129.

(10) Lipscomb, L. A.; Peek, M. E.; Zhou, F. X.; Bertrand, J. A.; VanDerveer, D.; Williams, L. D. Water Ring Structure at Dna Interfaces - Hydration and Dynamics of Dna Anthracycline Complexes. *Biochemistry* **1994**, *33*, 3649.

(11) Tu, C.; Rowlett, R. S.; Tripp, B. C.; Ferry, J. G.; Silverman, D. N. Chemical Rescue of Proton Transfer in Catalysis by Carbonic Anhydrases in the Beta- and Gamma-Class. *Biochemistry* **2002**, *41*, 15429.

(12) Soudackov, A. V.; Hammes-Schiffer, S. Probing Non-adiabaticity in the Proton-Coupled Electron Transfer Reaction Catalyzed by Soybean Lipoxygenase. *J. Phys. Chem. Lett.* **2014**, *5*, 3274.

(13) Li, P.; Soudackov, A. V.; Hammes-Schiffer, S. Fundamental Insights into Proton-Coupled Electron Transfer in Soybean Lipoxygenase from Quantum Mechanical/Molecular Mechanical Free Energy Simulations. *J. Am. Chem. Soc.* **2018**, *140*, 3068−3076.

(14) McEwan, M. J.; Phillips, L. F. *Chemistry of the Atmosphere*; Eward Arnold: London, 1975.

(15) Wayne, R. P. *Chemistry of the Atmosphere*; Clarendon Press: Oxford, 1994.

(16) Dietrick, S. M.; Pacheco, A. B.; Phatak, P.; Stevens, P. S.; Iyengar, S. S. Influence of Water on Anharmonicity, Stability, and Vibrational Energy Distribution of Hydrogen-Bonded Adducts in Atmospheric Reactions: Case Study of the OH + Isoprene Reaction

Intermediate Using Ab Initio Molecular Dynamics. *J. Phys. Chem. A* **2012**, *116*, 399.

(17) Iyengar, S. S. Dynamical Effects on Vibrational and Electronic Spectra of Hydroperoxyl Radical Water Clusters. *J. Chem. Phys.* **2005**, *123*, 084310.

(18) Schlegel, H. B.; Frisch, M. J. Computational bottlenecks in molecular orbital calculations. *Theoretical and Computational Models for Organic Chemistry*; Springer, 1991; pp 5−33.

(19) White, C. A.; Head-Gordon, M. Derivation and efficient implementation of the fast multipole method. *J. Chem. Phys.* **1994**, *101*, 6593.

(20) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. Achieving Linear Scaling for the Electronic Quantum Coulomb Problem. *J. Sci.* **1996**, *271*, 51.

(21) Ayala, P. Y.; Scuseria, G. E. Linear scaling second-order Moller−Plesset theory in the atomic orbital basis for large molecular systems. *J. Chem. Phys.* **1999**, *110*, 3660−3671.

(22) Iyengar, S. S.; Scuseria, G. E.; Savin, A. Bounding the Extrapolated Correlation Energy Using Pade Approximants. *Int. J. Quantum Chem.* **2000**, *79*, 222−234.

(23) Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **1982**, *21*, 467−488.

(24) Beck, M.; Jackle, A.; Worth, G. A.; Meyer, H.-D. The multiconfiguration time-dependent Hartree (MCTDH) method: a highly efficient algorithm for propagating wavepackets. *Phys. Rep.* **2000**, *324*, 1−105.

(25) Iyengar, S. S.; Kouri, D. J.; Hoffman, D. K. Particular and Homogeneous Solutions of Time-Independent Wavepacket Schrodinger Equations: Calculations Using a Subset of Eigenstates of Undamped or Damped Hamiltonians. *Theor. Chem. Acc.* **2000**, *104*, 471.

(26) Iyengar, S. S.; Jakowski, J. Quantum Wavepacket Ab Initio Molecular Dynamics: An Approach to Study Quantum Dynamics in Large Systems. *J. Chem. Phys.* **2005**, *122*, 114105.

(27) DeGregorio, N.; Iyengar, S. S. Adaptive dimensional decoupling for compression of quantum nuclear wavefunctions and efficient potential energy surface representations through tensor network decomposition. *J. Chem. Theory Comput.* **2019**, *15*, 2780−2796.

(28) Rabitz, H.; Aliş, Ö. F. General foundations of high dimensional model representations. *J. Math. Chem.* **1999**, *25*, 197−233.

(29) Manzhos, S.; Carrington, T., Jr. Using neural networks, optimized coordinates, and high-dimensional model representations to obtain a vinyl bromide potential surface. *J. Chem. Phys.* **2008**, *129*, 224104.

(30) Otto, F. Multi-layer Potfit: An accurate potential representation for efficient high-dimensional quantum dynamics. *J. Chem. Phys.* **2014**, *140*, 014106.

(31) 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.

(32) 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.

(33) Braams, B. J.; Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577.

(34) Yang, W. Direct Calculation of Electron Density in Density-Functional Theory. *Phys. Rev. Lett.* **1991**, *66*, 1438.

(35) 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.

(36) Hopkins, B. W.; Tschumper, G. S. Multicentred QM/QM Methods for Overlapping Model Systems. *Mol. Phys.* **2005**, *103*, 309.

(37) 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.

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

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

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

(41) 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.

(42) 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.

(43) 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.

(44) Beran, G. J. O. Modeling Polymorphic Molecular Crystals with Electronic Structure Theory. *Chem. Rev.* **2016**, *116*, 5567.

(45) Stoll, H. The correlation energy of crystalline silicon. *Chem. Phys. Lett.* **1992**, *191*, 548.

(46) Stoll, H. Correlation energy of diamond. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *46*, 6700.

(47) Stoll, H. On the correlation energy of graphite. *J. Chem. Phys.* **1992**, *97*, 8449.

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

(49) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701.

(50) 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.

(51) Huang, L.; Massa, L.; Karle, J. Kernel energy method: Application to DNA. *Biochemistry* **2005**, *44*, 16747.

(52) Wang, L.-W.; Zhao, Z.; Meza, J. Linear-Scaling Three-Dimensional Fragment method for Large-scale Electronic Structure Calculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 165113.

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

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

(55) Wen, S.; Nanda, K.; Huang, Y.; Beran, G. J. O. Practical quantum mechanics-based fragment methods for predicting molecular crystal properties. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7578.

(56) 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.

(57) Frankcombe, T. J.; Collins, M. A. Growing Fragmented Potentials for Gas-Surface Reactions: The Reaction between Hydrogen Atoms and Hydrogen-Terminated Silicon (111). *J. Phys. Chem. C* **2012**, *116*, 7793−7802.

(58) 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.

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

(60) Lange, A. W.; Voth, G. A. Multi-State Approach to Chemical Reactivity in Fragment Based Quantum Chemistry Calculations. *J. Chem. Theory Comput.* **2013**, *9*, 4018.

(61) 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.

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

(63) 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.

(64) 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.

(65) Herbert, J. M. Fantasy versus reality in fragment-based quantum chemistry. *J. Chem. Phys.* **2019**, *151*, 170901.

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

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

(68) Varandas, A. J. C.; Brown, F. B.; Mead, C. A.; Truhlar, D. G.; Blais, N. C. A double many-body expansion of the two lowest-energy potential surfaces and nonadiabatic coupling for $H_3$. *J. Chem. Phys.* **1987**, *86*, 6258−6269.

(69) Varandas, A. J. C.; Pais, A. A. C. C. A realistic double many-body expansion (DMBE) potential energy surface for ground-state $O_3$ from a multiproperty fit to ab initio calculations, and to experimental spectroscopic, inelastic scattering, and kinetic isotope thermal rate data. *Mol. Phys.* **1988**, *65*, 843.

(70) Lynch, G. C.; Steckler, R.; Schwenke, D. W.; Varandas, A. J. C.; Truhlar, D. G.; Garrett, B. C. Use of scaled external correlation, a double many-body expansion, and variational transition state theory to calibrate a potential energy surface for $FH_2$. *J. Chem. Phys.* **1991**, *94*, 7136−7149.

(71) 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.

(72) 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.

(73) Stoll, H.; Paulus, B.; Fulde, P. On the accuracy of correlation-energy expansions in terms of local increments. *J. Chem. Phys.* **2005**, *123*, 144108.

(74) 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.

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

(76) Hirata, S. Fast Electron-Correlation Methods for Molecular Crystals: an Application to the $\alpha$, $\beta(1)$, and $\beta(2)$ Modifications of Solid Formic Acid. *J. Chem. Phys.* **2008**, *129*, 204104.

(77) Beran, G. J. O.; Nanda, K. Predicting Organic Crystal Lattice Energies with Chemical Accuracy. *J. Phys. Chem. Lett.* **2010**, *1*, 3480−3487.

(78) 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.

(79) Bates, D. M.; Smith, J. R.; Janowski, T.; Tschumper, G. S. Development of a 3-body: many-body integrated fragmentation method for weakly bound clusters and application to water clusters $(H_2O)_{(n=3)−10,16,17}$. *J. Chem. Phys.* **2011**, *135*, 044123.

(80) Bygrave, P. J.; Allan, N. L.; Manby, F. R. The embedded many-body expansion for energetics of molecular crystals. *J. Chem. Phys.* **2012**, *137*, 164102.

(81) 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.

(82) 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.

(83) 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.

(84) Demerdash, O.; Mao, Y.; Liu, T.; Head-Gordon, M.; Head-Gordon, T. Assessing many-body contributions to intermolecular interactions of the AMOEBA force field using energy decomposition analysis of electronic structure calculations. *J. Chem. Phys.* **2017**, *147*, 161721.

(85) Yu, Q.; Bowman, J. M. Communication: VSCF/VCI vibrational spectroscopy of $H7 O 3 +$ and $H9 O 4 +$ using high-level, many-body potential energy surface and dipole moment surfaces. *J. Chem. Phys.* **2017**, *146*, 121102.

(86) Yao, K.; Herr, J. E.; Parkhill, J. The many-body expansion combined with neural networks. *J. Chem. Phys.* **2017**, *146*, 014106.

(87) Raghavachari, K.; Rohlfing, C. M. Bonding and stabilities of small silicon clusters: a theoretical study of Si7−Si10. *J. Chem. Phys.* **1988**, *89*, 2219−2234.

(88) Trucks, G. W.; Raghavachari, K.; Higashi, G. S.; Chabal, Y. J. Mechanism of HF etching of silicon surfaces: A theoretical understanding of hydrogen passivation. *Phys. Rev. Lett.* **1990**, *65*, 504.

(89) Grossman, J. C.; Mitas, L.; Raghavachari, K. Structure and stability of molecular carbon: importance of electron correlation. *Phys. Rev. Lett.* **1995**, *75*, 3870.

(90) Almlöf, J. Elimination of energy denominators in Møller-Plesset perturbation theory by a Laplace transform approach. *Chem. Phys. Lett.* **1991**, *181*, 319−320.

(91) Schütz, M. A new, fast, semi-direct implementation of linear scaling local coupled cluster theory. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3941−3947.

(92) Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.* **2003**, *118*, 8149−8160.

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

(94) Peverati, R.; Truhlar, D. G. The Quest for a Universal Density Functional: The Accuracy of Density Functionals Across a Broad Spectrum of Databases in Chemistry and Physics. *Philos. Trans. R. Soc., A* **2014**, *372*, 20120476.

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

(96) Komeiji, Y.; Mochizuki, Y.; Nakano, T. Three-Body Expansion and Generalized Dynamic Fragmentation Improve the Fragment Molecular Orbital-Based Molecular Dynamics (FMO-MD). *Chem. Phys. Lett.* **2010**, *484*, 380.

(97) Liu, J.; Zhu, T.; Wang, X.; He, X.; Zhang, J. Z. H. Quantum Fragment Based ab Initio Molecular Dynamics for Proteins. *J. Chem. Theory Comput.* **2015**, *11*, 5897.

(98) 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.

(99) 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.

(100) Collins, M. A. Can Systematic Molecular Fragmentation Be Applied to Direct Ab Initio Molecular Dynamics? *J. Phys. Chem. A* **2016**, *120*, 9281.

(101) Liu, P.; Li, W.; Kan, Z.; Sun, H.; Ma, J. Factor Analysis of Conformations and NMR Signals of Rotaxanes: AIMD and Polarizable MD Simulations. *J. Phys. Chem. A* **2016**, *120*, 490.

(102) Li, J.; Haycraft, C.; Iyengar, S. S. Hybrid extended Lagrangian, post-Hartree-Fock Born-Oppenheimer ab initio molecular dynamics using fragment-based electronic structure. *J. Chem. Theory Comput.* **2016**, *12*, 2493.

(103) 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.

(104) 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.

(105) Liu, J.; Rana, B.; Liu, K.-Y.; Herbert, J. M. Variational Formulation of the Generalized Many-Body Expansion with Self-Consistent Charge Embedding: Simple and Correct Analytic Energy Gradient for Fragment-Based ab Initio Molecular Dynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 3877.

(106) 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.

(107) 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.

(108) König, C.; Christiansen, O. Linear-scaling generation of potential energy surfaces using a double incremental expansion. *J. Chem. Phys.* **2016**, *145*, 064105.

(109) Hua, S.; Hua, W.; Li, S. An efficient implementation of the generalized energy-based fragmentation approach for general large molecules. *J. Phys. Chem. A* **2010**, *114*, 8126−8134.

(110) He, X.; Zhu, T.; Wang, X.; Liu, J.; Zhang, J. Z. H. Fragment quantum mechanical calculation of proteins and its applications. *Acc. Chem. Res.* **2014**, *47*, 2748−2757.

(111) Liu, J.; Herbert, J. M. Pair−pair approximation to the generalized many-body expansion: an alternative to the four-body expansion for ab initio prediction of protein energetics via molecular fragmentation. *J. Chem. Theory Comput.* **2016**, *12*, 572−584.

(112) 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.

(113) 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*, 1887.

(114) Ricard, T. C.; Kumar, A.; Iyengar, S. S. Embedded, graph-theoretically defined many-body approximations for wavefunction-in-DFT and DFT-in-DFT: applications to gas- and condensed-phase AIMD, and potential surfaces for quantum nuclear effects. *Int. J. Quantum Chem.* **2020**, *120*, No. e26244.

(115) 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.

(116) 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.

(117) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ oxidative addition. *J. Phys. Chem.* **1996**, *100*, 19357.

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

(119) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. The ONIOM Method: Its Foundation and Applications to Metal-loenzymes and Photobiology. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 327.

(120) 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.

(121) Schult, D. A.; Swart, P. Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conferences (SciPy 2008)*, 2008; pp 11−16.

(122) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261−272.

(123) Sontising, W.; Beran, G. J. O. Theoretical assessment of the structure and stability of the $\lambda$ phase of nitrogen. *Phys. Rev. Mater.* **2019**, *3*, 095002.

(124) Abusalbi, N.; Kouri, D. J.; Lopez, V.; Babamov, V. K.; Marcus, R. A. Comparison of multichannel and two-state calculations for H-atom transfer between two nearly degenerate states. *Chem. Phys. Lett.* **1984**, *103*, 458−462.

(125) 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, New York, 1996; p 411.

(126) Yarkony, D. R. Diabolical Conical Intersections. *Rev. Mod. Phys.* **1996**, *68*, 985.

(127) Schmitt, U. W.; Voth, G. A. Multistate empirical valence bond model for proton transport in water. *J. Phys. Chem. B* **1998**, *102*, 5547−5551.

(128) Kendrick, B. K.; Mead, C. A.; Truhlar, D. G. Properties of Nonadiabatic Couplings and the Generalized Born Oppenheimer Approximation. *Chem. Phys.* **2002**, *277*, 31.

(129) 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.

(130) Worth, G. A.; Cederbaum, L. S. Beyond Born-Oppenheimer: Molecular Dynamics Through a Conical Intersection. *Annu. Rev. Phys. Chem.* **2004**, *55*, 127.

(131) Jasper, A. W.; Zhu, C.; Nangia, S.; Truhlar, D. G. Introductory Lecture: Nonadiabatic Effects in Chemical Dynamics. *Faraday Discuss.* **2004**, *127*, 1.

(132) Baer, M. *Beyond Born-Oppenheimer: Conical Intersections and Electronic Nonadiabatic Coupling Terms*; Wiley: New York, 2006.

(133) Hanna, G.; Kapral, R. Nonadiabatic Dynamics of Condensed Phase Rate Processes. *Acc. Chem. Res.* **2006**, *39*, 21.

(134) Park, K.; Lin, W.; Paesani, F. A refined MS-EVB model for proton transport in aqueous environments. *J. Phys. Chem. B* **2012**, *116*, 343−352.

(135) Mulliken, R. S. The interaction of electron donors and acceptors. *J. Chim. Phys.* **1964**, *61*, 20.

(136) 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.

(137) 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.

(138) Borgis, D.; Staib, A. A semiempirical quantum polarization model for water. *Chem. Phys. Lett.* **1995**, *238*, 187.

(139) Vuilleumier, R.; Borgis, D. Quantum Dynamics of an Excess Proton in Water Using an Extended Empirical Valence-Bond Hamiltonian. *J. Phys. Chem. B* **1998**, *102*, 4261−4264.

(140) Vuilleumier, R.; Borgis, D. An Extended Empirical Valence Bond Model for Describing Proton Mobility in Water. *Isr. J. Chem.* **1999**, *39*, 457.

(141) Vuilleumier, R.; Borgis, D. Transport and Spectroscopy of the Hydrated Proton: A Molecular Dynamics Study. *J. Chem. Phys.* **1999**, *111*, 4251.

(142) Lefohn, A. E.; Ovchinnikov, M.; Voth, G. A. A Multistate Empirical Valence Bond Approach to a Polarizable and Flexible Water Model†. *J. Phys. Chem. B* **2001**, *105*, 6628.

(143) Cuma, M.; Schmitt, U. W.; Voth, G. A. A Multi-State Empirical Valence Bond Model for Weak Acid Dissociation in Aqueous Solution. *J. Phys. Chem. A* **2001**, *105*, 2814−2823.

(144) Day, T. J. F.; Soudackov, A. V.; Cuma, 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.

(145) 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.

(146) Petersen, M. K.; Iyengar, S. S.; Day, T. J. F.; Voth, G. A. The hydrated proton at the water liquid/vapor interface. *J. Phys. Chem. B* **2004**, *108*, 14804−14806.

(147) Song, L.; Gao, J. On the Construction of Diabatic and Adiabatic Potential Energy Surfaces Based on Ab Initio Valence Bond Theory. *J. Phys. Chem. A* **2008**, *112*, 12925−12935.

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

(149) Adams, C. C.; Franzosa, R. D. *Introduction to Topology: Pure and Applied*; Pearson Prentice Hall Upper: Saddle River, 2008.

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

(151) Moon, J. W.; Moser, L. On cliques in graphs. *Isr. J. Math.* **1965**, *3*, 23−28.

(152) Balas, E.; Yu, C. S. Finding a maximum clique in an arbitrary graph. *SIAM J. Comput.* **1986**, *15*, 1054−1068.

(153) Alış, Ö. F.; Rabitz, H. Efficient implementation of high dimensional model representations. *J. Math. Chem.* **2001**, *29*, 127−142.

(154) Jäckle, A.; Meyer, H. D. Product representation of potential energy surfaces. *J. Chem. Phys.* **1996**, *104*, 7974−7984.

(155) Jäckle, A.; Meyer, H.-D. Product representation of potential energy surfaces. II. *J. Chem. Phys.* **1998**, *109*, 3772−3779.

(156) Conte, R.; Qu, C.; Bowman, J. M. Permutationally invariant fitting of many-body, non-covalent interactions with application to three-body methane−water−water. *J. Chem. Theory Comput.* **2015**, *11*, 1631−1638.

(157) Lorentz, G.; Golitschek, M.; Makovoz, Y. *Constructive Approximation*; Springer: New York, 1996.

(158) Girosi, F.; Poggio, T. Representation properties of networks: Kolmogorov's theorem is irrelevant. *Neurocomputing* **1989**, *1*, 465−469.

(159) Sobol', I. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 86−112.

(160) 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.

(161) Collins, M. A. Molecular potential energy surfaces constructed from interpolation of systematic fragment surfaces. *J. Chem. Phys.* **2007**, *127*, 024104.

(162) Jakowski, J.; Sumner, I.; Iyengar, S. S. Computational Improvements to Quantum Wavepacket Ab Initio Molecular Dynamics Using a Potential-Adapted, Time-Dependent Deterministic Sampling Technique. *J. Chem. Theory Comput.* **2006**, *2*, 1203.

(163) 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.

(164) 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.

(165) Avila, G.; Carrington, T. Nonproduct quadrature grids for solving the vibrational Schrödinger equation. *J. Chem. Phys.* **2009**, *131*, 174103.

(166) Flynn, S. W.; Mandelshtam, V. A. Sampling general distributions with quasi-regular grids: Application to the vibrational spectra calculations. *J. Chem. Phys.* **2019**, *151*, 241105.

(167) 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.

(168) Du, Q.; Faber, V.; Gunzburger, M. Centroidal Voronoi tessellations: Applications and algorithms. *SIAM Rev.* **1999**, *41*, 637−676.

(169) Tournois, J.; Alliez, P.; Devillers, O. 2D centroidal Voronoi tessellations with constraints. *Numer. Math.* **2010**, *3*, 212−222.

(170) Urschel, J. C. On the characterization and uniqueness of centroidal Voronoi tessellations. *SIAM J. Numer. Anal.* **2017**, *55*, 1525−1547.

(171) Coffey, T. M.; Wyatt, R. E.; Schieve, W. C. Quantum Trajectories from Kinematic Considerations. *J. Phys. A: Math. Theor.* **2010**, *43*, 335301.

(172) Bowyer, A. Computing dirichlet tessellations. *Comp. J.* **1981**, *24*, 162.

(173) Lee, D.-T.; Preparata, F. P. Computational Geometry-A Survey. *IEEE Trans. Comput.* **1984**, *33*, 1072−1101.

(174) Aurenhammer, F. Voronoi diagrams-a survey of a fundamental geometric data structure. *ACM Comput. Surv.* **1991**, *23*, 345.

(175) Okabe, A.; Boots, B.; Sugihara, K.; Chiu, S. N. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*; John Wiley and Sons, 2000.

(176) Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129−137.

(177) Warshel, A.; Papazyan, A.; Kollman, P. On low-barrier hydrogen bonds and enzyme catalysis. *Science* **1995**, *269*, 102.

(178) Chatfield, D. C.; Brooks, B. R. HIV-1 Protease Cleavage Mechanism Elucidated with Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **1995**, *117*, 5561.

(179) Gerlt, J. A.; Kreevoy, M. M.; Cleland, W. W.; Frey, P. A. Understanding Enzymic Catalysis: The Importance of Short, Strong Hydrogen Bonds. *Chem. Biol.* **1997**, *4*, 259.

(180) Azzouz, H.; Borgis, D. A quantum molecular-dynamics study of proton-transfer reactions along asymmetrical H bonds in solution. *J. Chem. Phys.* **1993**, *98*, 7361.

(181) Cleland, W.; Kreevoy, M. Low-barrier hydrogen bonds and enzymic catalysis. *Science* **1994**, *264*, 1887.

(182) Nagle, J. F.; Morowitz, H. J. Molecular Mechanisms for Proton Transport in Membranes. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 298.

(183) Baciou, L.; Michel, H. Interruption of the water chain in the reaction center from Rhodobacter sphaeroides reduces the rates of the proton uptake and of the second electron transfer to QB. *Biochemistry* **1995**, *34*, 7967−7972.

(184) Guo, H.; Barnard, A. S. Proton transfer in the hydrogen-bonded chains of lepidocrocite: a computational study. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17864.

(185) Domene, C.; Sansom, M. S. P. Potassium channel, ions, and water: simulation studies based on the high resolution X-ray structure of KcsA. *Biophys. J.* **2003**, *85*, 2787−2800.

(186) Ye, Y.-S.; Rick, J.; Hwang, B.-J. Water soluble polymers as proton exchange membranes for fuel cells. *Polymers* **2012**, *4*, 913−963.

(187) Tuckerman, M. E.; Ungar, P. J.; von Rosenvinge, T.; Klein, M. L. Ab Initio Molecular Dynamics Simulations. *J. Phys. Chem.* **1996**, *100*, 12878.

(188) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. Ab Initio Molecular Dynamics Simulation of the Solvation and Transport of $H_3O^+$ and $OH^-$ Ions in Water. *J. Phys. Chem.* **1995**, *99*, 5749.

(189) Schmitt, U. W.; Voth, G. A. The computer simulation of proton transport in water. *J. Chem. Phys.* **1999**, *111*, 9361−9381.

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

(191) Sorensen, D. C. Implicit Application of Polynomial Filters in a K-Step Arnoldi Method. *SIAM J. Matrix Anal. Appl.* **1992**, *13*, 357.

(192) Parlett, B. N.; Saad, Y. Complex shift and invert strategies for real matrices. *Lin. Algebra Appl.* **1987**, *88−89*, 575−595.

(193) Headrick, J. M.; Diken, E. G.; Walters, R. S.; Hammer, N. I.; Christie, R. A.; Cui, J.; Myshakin, E. M.; Duncan, M. A.; Johnson, M. A.; Jordan, K. Spectral Signatures of Hydrated Proton Vibrations in Water Clusters. *Science* **2005**, *308*, 1765.

(194) Diken, E. G.; Headrick, J. M.; Roscioli, J. R.; Bopp, J. C.; Johnson, M. A.; McCoy, A. B. Fundamental Excitations of the Shared Proton in the $H_3O_2^-$ and $H_5O_2^+$ Complexes. *J. Phys. Chem. A* **2005**, *109*, 1487.

(195) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry. *J. Phys. Chem. A* **2001**, *105*, 293−307.

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

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

(198) Brorsen, K. R.; Minezawa, N.; Xu, F.; Windus, T. L.; Gordon, M. S. Fragment Molecular Orbital Molecular Dynamics with the Fully Analytic Energy Gradient. *J. Chem. Theory Comput.* **2012**, *8*, 5008.

(199) Brorsen, K. R.; Zahariev, F.; Nakata, H.; Fedorov, D. G.; Gordon, M. S. Analytic Gradient for Density Functional Theory Based on the Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2014**, *10*, 5297.

(200) Hoffman, D. K.; Nayar, N.; Sharafeddin, O. A.; Kouri, D. J. Analytic banded approximation for the discretized free propagator. *J. Phys. Chem.* **1991**, *95*, 8299−8305.

(201) Kouri, D. J.; Huang, Y.; Hoffman, D. K. Iterated real-time path integral evaluation using a distributed approximating functional propagator and average-case complexity integration. *Phys. Rev. Lett.* **1995**, *75*, 49.

(202) Iyengar, S. S. Ab Initio Dynamics with Wave-Packets and Density Matrices. *Theor. Chem. Acc.* **2006**, *116*, 326.

(203) Pacheco, A. B.; Iyengar, S. S. Multi-Stage Ab-Initio Quantum Wavepacket Dynamics for Electronic Structure and Dynamics in Open Systems: Momentum Representation, Coupled Electron Nuclear Dynamics and External Fields. *J. Chem. Phys.* **2011**, *134*, 074107.

(204) Korevaar, J. Pansions and the Theory of Fourier Transforms. *Trans. Am. Math. Soc.* **1959**, *91*, 53.

(205) Yu, S.; Zhao, S.; Wei, G. W. Local spectral time splitting method for first- and second-order partial differential equations. *J. Comput. Phys.* **2005**, *206*, 727−780.

(206) *Handbook of Mathematical Functions*; Abramowitz, M., Stegun, I. A., Eds.; U.S. GPO: Washington, D.C., 1964.

(207) Hoffman, D. K.; Arnold, M.; Kouri, D. J. Properties of the optimum distributed approximating function class propagator for discretized and continuous wave packet propagations. *J. Phys. Chem.* **1992**, *96*, 6539−6545.

(208) Grossmann, A.; Morlet, J. Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape. *SIAM J. Math. Anal.* **1984**, *15*, 723−736.

(209) Strang, G. Wavelets and Dilation Equations: A Brief Introduction. *SIAM Rev.* **1989**, *31*, 614−627.

(210) Daubechies, I. *Ten Lectures on Wavelets*; SIAM, 1992.

(211) Strang, G.; Nguyen, T. Q. *Wavelets and Filter Banks*; SIAM, 1996.

(212) Billing, G. D. Quantum dressed classical mechanics. *J. Chem. Phys.* **2001**, *114*, 6641−6653.

(213) Ben-Nun, M.; Quenneville, J.; Martínez, T. J. Ab Initio Multiple Spawning: Photochemistry from First Principles Quantum Molecular Dynamics. *J. Phys. Chem. A* **2000**, *104*, 5161−5175.

(214) Heller, E. J. Time-dependent approach to semiclassical dynamics. *J. Chem. Phys.* **1975**, *62*, 1544−1555.