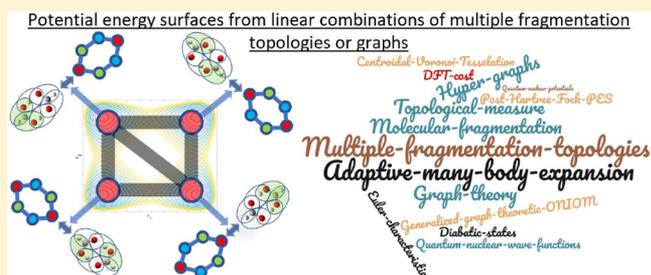


# Fragment-Based Electronic Structure for Potential Energy Surfaces Using a Superposition of Fragmentation Topologies

Anup Kumar and Srinivasan S. Iyengar\*<sup>✉</sup>

Department of Chemistry and Department of Physics, Indiana University, 800 East Kirkwood Avenue, Bloomington, Indiana-47405, United States

**ABSTRACT:** We present a new approach for adaptive molecular fragmentation. Here multiple fragmentation protocols, or fragmentation topologies, are combined to efficiently and accurately construct potential energy surfaces that are in agreement with post-Hartree–Fock levels of electronic structure theories at density functional theory (DFT) cost. We benchmark the method through evaluation of quantum nuclear effects in a set of protonated water clusters that are known to display significant quantum effects. In such systems, the straightforward use of molecular fragmentation is hindered by the fact that the most appropriate fragmentation strategy changes as a function of nuclear degrees of freedom. Our approach uses a multilayered hypergraph formalism to decompose the potential energy surface, where, at the very top layer, a tessellation of the potential surface yields a set of independent, but correlated, graphical nodes or vertices; each node represents a different protocol to fragment the molecular system. Correlation between the nodes appears as edges and faces in the graph at the top layer and allows the overall potential surface to be represented as a superposition of multiple fragmentation topologies with the coefficients for the superposition arising from a Hamiltonian formalism that is reminiscent of nonadiabatic dynamics. This allows for a natural interpretation of the individual molecular fragmentation topologies as diabatic or valence-bond-type states which we exploit in our formalism. As stated, the method is demonstrated for protonated water clusters where we are able to obtain potentials surfaces in agreement with post-Hartree–Fock methods at DFT cost.



## I. INTRODUCTION

In a recent set of publications,<sup>1–5</sup> we have shown how both extended Lagrangian and Born–Oppenheimer-based ab initio molecular dynamics simulations can be performed at accuracy comparable to CCSD and MP2 levels of theory with density functional theory (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 efforts commensurate with much smaller basis set sizes. In ref 5, 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 polypeptides systems with 100+ atoms. Specifically, (a) in refs 1 and 2, we have used this approach to compute Born–Oppenheimer dynamics trajectories that agree with MP2-based AIMD at DFT expense. (b) In refs 2 and 3, we have introduced extended-Lagrangian versions of the same that are in agreement with MP2- and CCSD-based Born–Oppenheimer molecular dynamics; and hence, for the first time, in refs 2 and 3, we presented Car–Parrinello-style dynamics, but with CCSD accuracy. (c) In refs 1 and 2, we have shown that potential surfaces in agreement with CCSD(T) can be constructed for protonated water clusters with computational effort that scales as DFT. (d) In ref 5, we have constructed an efficient approach for on-the-fly basis set extrapolation in extended Lagrangian and Born–Oppenheimer

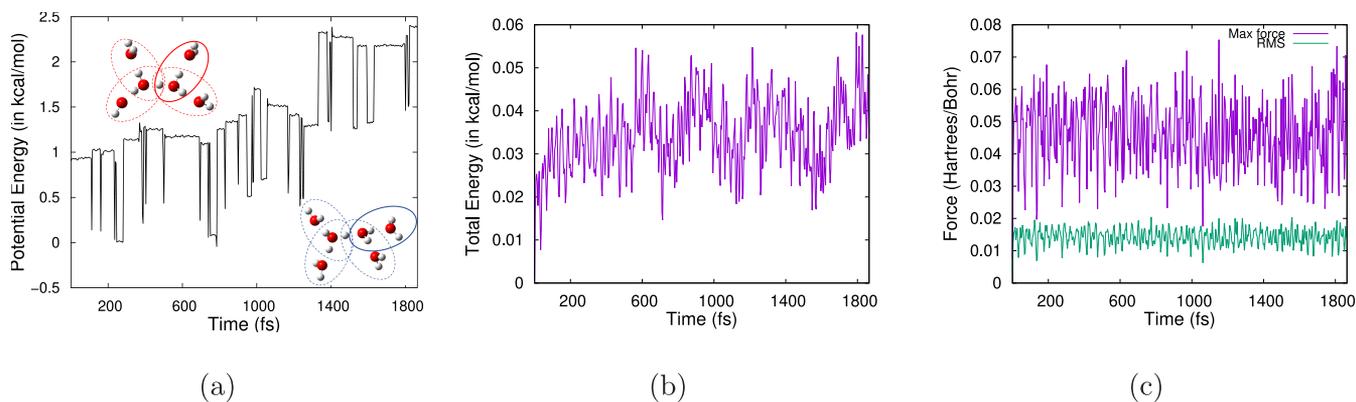
dynamics. The Born–Oppenheimer version of this approach is called frag-BOMD, whereas the extended-Lagrangian generalization is called atom-centered density matrix propagation with post-Hartree–Fock accuracy (ADMP-pHF) since it derives from the ADMP<sup>6–9</sup> extended Lagrangian. The approach described above uses molecular fragmentation,<sup>10–38</sup> but the method in refs 1–3 is a set-theoretic, inclusion–exclusion principle<sup>39</sup> generalization of ONIOM<sup>16</sup> with overlapping “model” systems that may cover the entire domain of the “real” system. In refs 4 and 5, the approach is adapted using graph-theoretic methods to provide improved computational efficiency and allows the construction of a “coarse-grained” version of AIMD.

These studies<sup>1–5</sup> are closely related to several other methods in the literature.<sup>11,21,22,27–30,40,41</sup> Although the formalism in refs 1–5 is derived from ONIOM,<sup>12</sup> it has connections to the multicentered QM:QM formalism,<sup>21,42</sup> the molecular tailoring approach (MTA),<sup>27,40</sup> the ONIOM-XO method,<sup>28</sup> and the molecules-in-molecules (MIM) methodology.<sup>11,29,30,41,43</sup> There are several fragmentation methods<sup>13,19,25,35,44–52</sup> available, but the approaches in refs 1–4, 11, 21, 28, 29, 40, and 41 include long-range electronic effects through a full-system low-level calculation, much in the same vein as ONIOM.<sup>12</sup> We have noted in ref 4 that the approach studied here is also

Received: June 20, 2019

Published: September 26, 2019



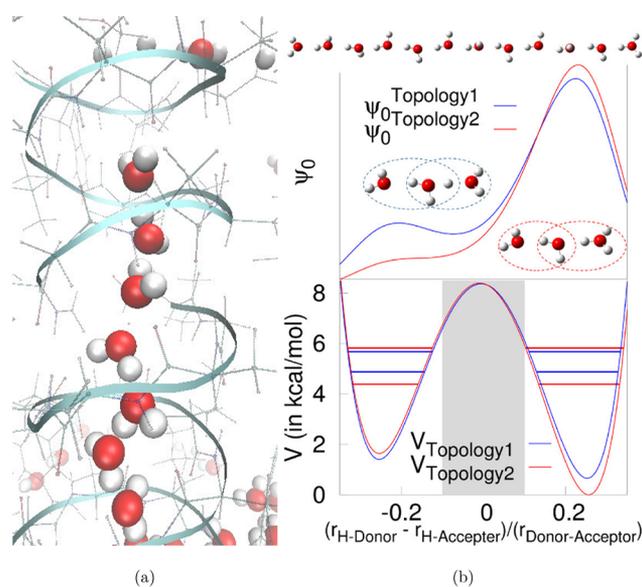


**Figure 1.** Evolution of electronic potential energy during a fragment-based BOMD calculations performed at the CCSD/6-31+g(d,p):B3LYP/6-31+g(d,p) level of theory for the solvated Zundel cation. The sudden hops in panel a are due to changes in fragment description, two of which are presented in panel a with ellipses describing fragments. Note that the figure with ellipses marked in blue has a different fragment description as compared to that in red. However, when the total energy is simply shifted by the extent of hop (b), the total energy is quite well behaved with deviations of the order of a few hundredths of a kcal/mol. In fact, from panel b, the RMS deviation in total energy is 0.009 kcal/mol and the associated drift in total energy is 0.019 kcal/mol that are well within acceptable range for AIMD simulations. Similarly the forces (maximum and RMS forces presented in panel c) are also well-behaved.

closely related to many-body expansions<sup>23,32,53</sup> and double many-body expansions.<sup>53</sup>

Despite the progress highlighted above, critical challenges remain. For example, the fragmentation approach is a dynamical phenomenon that in general changes with time during AIMD and in general with change in nuclear configuration. Thus, in addition to having a role in AIMD, the dynamical nature of the molecular fragmentation would be of critical importance in constructing potential energy surfaces to compute quantum nuclear effects. This is significant for studies involving protonated and hydroxide-rich water clusters, where the excess charge hops and may be delocalized through a dynamical network of water molecules. Similarly, one can also expect that the dynamical nature of molecular fragmentation is critical in hydrogen transfer reactions<sup>60</sup> where nuclear quantization<sup>61</sup> has been known to have a critical role.

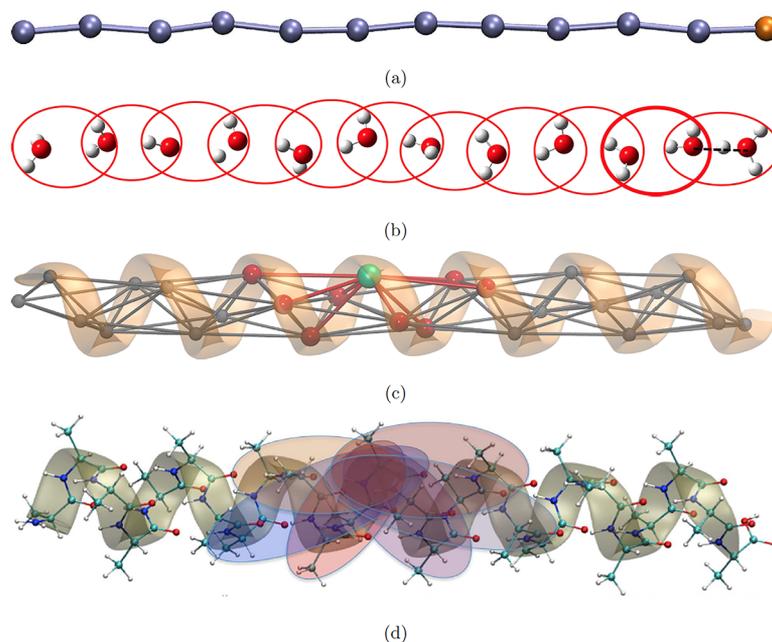
In Figure 1, we present an illustration of the effect of the change in fragmentation in classical dynamics where the evolution of potential energy for a solvated Zundel cation system is presented from a fragmentation-based Born–Oppenheimer molecular dynamics simulation performed at the CCSD/6-31+g(d,p):B3LYP/6-31+g(d,p) level of theory. The sudden changes in potential energy are time steps where the fragment topology changes in concert with geometric changes. Two of the many fragmentation topologies are shown in Figure 1. While these hops lead to discontinuities in the total energy, the shifted energies and the forces are well-behaved and do not show significant deviations associated with change in topology. This is, however, not the case for quantum nuclear effects, which are generally nonlocal and are described in Figure 2. In Figure 2a we show a water–wire inside the Gramicidin-A proton channel.<sup>54,56,57</sup> In Figure 2b, we provide an illustration of the effect of change in fragment topology during a reduced dimensional potential surface calculation to account for nuclear quantum effects along the water–wire in the Gramicidin-A. The difference between eigenenergies (shown using horizontal lines on the potential surfaces in Figure 2b) of ground and first excited states (i.e., the excitation energy) obtained using two different fragmentation topologies is 0.80 and 1.44 kcal/mol, respectively. The differences in the



**Figure 2.** (a) Water–wire system inside the Gramicidin-A ion channel.<sup>54–57</sup> Such protonated water–wire systems are known for their quantum nuclear behavior arising from the shared proton delocalization.<sup>58,59</sup> Correspondingly, a potential surface for the shared (excess) proton in a water–wire system (shown on top in panel b) is shown in the bottom of panel b. The fact that the fragments shown in red have different identities compared to those in blue leads to two different surfaces with correspondingly different eigenenergies and quantum nuclear eigenfunctions, where the zero-point and first excited state levels are marked.

respective ground state wave functions are also noted from the top panel of Figure 2b. In this work, we introduce a method to use a set of fragmentation protocols, such as those in Figure 2, together to obtain efficient, high-quality potential surfaces.

This work is organized as follows: In section II, the theoretical aspects of the fragmentation-based methods used in this study are discussed, subsection II.A presents formal constructs for the use of multiple fragmentation topologies for potential surface calculations; a general algorithm is discussed in subsection II.B. Section III provides details regarding the



**Figure 3.** Illustration of the graph-based and set-theoretic fragmentation for water–wires (a and b) and polyaniline (c and d).

simulations performed on a benchmark protonated water–wire system similar to those in Figure 2. Subsections III.B and III.C include a detailed analysis of results obtained from including a superposition of fragmentation topologies in a continuous manner, and hence the method in section III.B is referred to here as “continuous topology morphing” and abbreviated as CTM. We consider the distinct cases of topology morphing where the morphing points, or seam of intersection between fragmentation protocols, is known exactly or is not known exactly. These discussions are critical toward generalization of the method presented to higher dimensions. Subsection III.D shows the interpolation for multidimensional potential surfaces utilized to capture quantum nuclear effects of coupled proton motions. Finally conclusions are given in section IV.

## II. MULTICONFIGURATIONAL FRAGMENTATION FOR MOLECULAR POTENTIAL ENERGY SURFACES: A STEP TOWARD REPRESENTING MOLECULAR CONFIGURATIONS AND POTENTIAL ENERGY SURFACES TOGETHER AS HYPERGRAPHS

In Figure 3a,c, we present a coarse-grained (CG) simplex decomposition<sup>62</sup> (or graphical representation) of the molecular framework defined for the systems in Figure 3b,d. Each vertex in Figure 3a,c represents a CG monomer, which for the

water–wire would either be one water molecule or a hydronium ion (see Figure 3a), and for the polypeptide it is a single amino acid fragment; for a general molecular system this would represent a chosen fragment. The connected edges shown in Figure 3a,c represent dimer fragments as seen in Figure 3b,d. Higher order simplex components of the graph that include trimer and tetrameric fragments and so on, are defined as seen in Figure 3c, and this constitutes the many-body, extended, coarse-graining method described in ref 4 through simplex decomposition and used for post-Hartree–Fock accuracy and in ref 5 for basis set extrapolation. Here, Figure 3 is provided for illustrative purposes to focus our discussion on potential energy surfaces obtained from multiple such graphical decompositions.

The molecular fragmentation in Figures 3b,d is algebraically translated in refs 1–3 using the set-theoretic principle of inclusion–exclusion,<sup>39</sup> where the molecular fragment depicted using edges is treated as sets and the vertices turn out to be overlapping regions between sets (see Figure 3b,d). These geometric parameters carry energy corrections to the overall system energy, and the associated inclusion–exclusion principle based generalization of the ONIOM method presented in refs 1–3 has the mathematical form

$$E^{\text{PIE-ONIOM}} \approx E^{\text{level},0}(0) + \sum_{i=1}^n \Delta E(i) - \sum_{1 \leq i < j \leq n} \Delta E(i \cap j) + \sum_{1 \leq i < j < k \leq n} \Delta E(i \cap j \cap k) - \dots + (-1)^{n-1} \sum_{1 < \dots < i < \dots < n} \Delta E(1 \cap \dots \cap i \cap \dots \cap n) \quad (1)$$

where  $E^{\text{level},0}(0)$  is the energy corresponding to the full system in Figure 3 at some predefined lower level of theory, whereas

$$\Delta E(i) = E^{\text{level},1}(i) - E^{\text{level},0}(i) \quad (2)$$

represents a correction term, as in ONIOM,<sup>16</sup> arising from each set in Figure 3b, represented using the indices  $i$  and  $j$ , that are in turn represented as edges in Figure 3a. Overlapping

fragments between sets are represented as  $i \cap j$  and so on. Thus, Figure 3b represents an hierarchy of fragments used to compute energies and forces.

In light of the decomposition shown in Figure 3c, a more sophisticated and computationally efficient<sup>4</sup> treatment uses the language of hypergraphs<sup>63,64</sup> and simplicial complexes.<sup>63,64</sup> Basically, a local complex is defined around each node, as

depicted in Figure 3c, and this local complex prescribes the level of local bonded and nonbonded interactions required to be captured on the basis of our approximation. In the discussion below, we assume this local complex to be of rank- $R$  and constant for all nodes, although this is not necessary. Figure 3a may be visualized as resulting from the local complex in Figure 3c floating from one node to the other. Such a graphical structure is called a simplicial complex, which is a special case of a hypergraph. Given such a network, it is always possible to define an invariant property of such a graph known as the Euler characteristic,<sup>63,64</sup> defined as

$$\begin{aligned}\chi &= \eta_0 - \eta_1 + \eta_2 - \dots + (-1)^r \eta_r + \dots + (-1)^R \eta_R \\ &= \sum_{r=0}^R (-1)^r \eta_r\end{aligned}\quad (3)$$

where  $\eta_r$  is the number of geometric entities (or simplexes) of rank- $r$ ; that is,  $\eta_0$  is the number of nodes,  $\eta_1$  is the number of edges,  $\eta_2$  is the number of faces, and so on. (Note: Geometric entities within a graph, nodes, edges, triangles, and tetrahedrons, are all finite-rank simplexes.<sup>62,65–70</sup> Thus, a node is a rank-0 simplex, an edge is a rank-1 simplex, and so on.) Here, we replace the appearance of each rank- $r$ , numbered using the index  $\alpha$ , by an energy correction analogous to eq 2, that is,  $\Delta E(\alpha, r)$ . Correspondingly the graph-theoretic energy expression,<sup>4</sup> which is analogous to eq 1, and consistent with the Euler characteristic in eq 3 may be written as

$$E^{\text{graph-theoretic}} = E^{\text{level},0} + \sum_{r=0}^R (-1)^r \left\{ \sum_{\alpha}^{r-\text{rank}} \Delta E(\alpha, r) \left[ \sum_{m=r}^R (-1)^m p_{\alpha}^{r,m} \right] \right\}\quad (4)$$

As noted above, “ $\alpha$ ” represents a simplex<sup>65–68</sup> of rank- $r$  and  $R$  is the largest simplex rank considered for electronic structure treatment. The curly bracketed term, {...}, refers to the energetic replacement for all rank- $r$  simplexes, that is,  $\eta_r$ , in eq 3:

$$\eta_r \rightarrow \left\{ \sum_{\alpha}^{r-\text{rank}} \Delta E(\alpha, r) \left[ \sum_{m=r}^R (-1)^m p_{\alpha}^{r,m} \right] \right\}\quad (5)$$

The square bracketed term contains an overcounting correction,<sup>4,5</sup> where  $p_{\alpha}^{r,m}$  is the number of times the  $\alpha$ th rank- $r$  simplex appears in all rank- $m$  simplexes ( $m > r$ ). Thus, the right side of eq 4 may be interpreted as the sum of a reference energy  $E^{\text{level},0}$  and a Euler characteristic-like energy correction that is now computed by constructing a measure on each simplex as depicted in eq 5. The quantity  $\Delta E(\alpha, r)$ , as noted above, is a difference in energy, analogous to  $\Delta E(i)$  in eq 2, that is, the difference in higher and lower level energies for the  $\alpha$ th rank- $r$  simplexes.

With respect to constructing electronic structure approximations using eq 4, there are two specific control parameters in eq 4. One pertains to the choice of “ $R$ ”, and the other pertains to the spatial envelope that assembles the family of edges or two-body interactions. The latter pertains to the neighborhood used to determine the set of  $\{\alpha\}$  for each “ $r$ ” in eq 4. (Note that once the vertices and edges are defined, the graph is completely defined and the expression in eq 4 can then be truncated at any order.) In this work,  $R$ , the maximum simplex rank, is chosen as one and only vertices and edges are

included in the fragmentation work. In this manner, the potential energy of a system, for a given configuration or a family of configurations, can be computed using either the set-theoretic (given by eq 1) or graph-theoretic (given by eq 4) approach. The geometric network implementation in Figure 3a allows a much more computationally efficient implementation than in eq 1 as discussed in ref 4. Nevertheless, as described earlier, the fragmentation energy will explicitly depend on the fragmentation protocol. For the case of Figure 3b, the fragmentation energy, eq 1, depends on the definition of the sets,  $\{\{i\}, \{i \cap j\}, \{i \cap j \cap k\}, \dots\}$ , and for the case in Figure 3a,c, the fragmentation energy, eq 4, depends on the adjacency matrix that defines the graph,  $\mathcal{G}$ , represented in Figure 3a,c. In either case, dynamical effects and potential energy surface calculations will require one to be able to consider multiple fragmentation protocols simultaneously.

**II.A. Formal Constructs for a Formalism That Uses Multiple Graphical Networks.** We begin by rewriting eqs 1 and 4 as a map between (i) a molecular geometry and (ii) a set of local bonded and nonbonded connectivities depicted by  $\{\{i\}, \{i \cap j\}, \{i \cap j \cap k\}, \dots\}$  or  $\mathcal{G}$  and an associated algebraic function,  $E^{\text{PIE-ONIOM}}$  or  $E^{\text{graph-theoretic}}$ . Specifically, the right sides of eqs 1 and 4 are completely determined by the molecular coordinates of the system and the chosen molecular fragmentation protocol; that is,

$$(\mathbf{R}, \tau) \equiv (\mathbf{R}, \{\{i\}, \{i \cap j\}, \{i \cap j \cap k\}, \dots\}) \equiv (\mathbf{R}, \mathcal{G})\quad (6)$$

Here  $\mathbf{R}$  represents the system coordinates, the sets  $\{i\}, \{i \cap j\}$ , and so on are together described as  $\tau$  and, similarly, for the graph  $\mathcal{G}$ , that is,

$$\tau \equiv \{\{i\}, \{i \cap j\}, \{i \cap j \cap k\}, \dots\} \equiv \mathcal{G}\quad (7)$$

and represents the molecular fragmentation or graph-theoretic decomposition protocol. Equation 6 also defines a “truncated” topological space.<sup>71–75</sup> The prescriptions  $\tau$ ,  $\mathcal{G}$ , and  $\{\{i\}, \{i \cap j\}, \{i \cap j \cap k\}, \dots\}$  imply a connectedness in the coordinate space,  $\mathbf{R}$ , by defining critical bonded and nonbonded interactions; this so-called connectedness is the essence of the nature in which electronic structure calculations are to be performed to obtain eqs 1 and 4. In this sense eqs 1 and 4 define a map:

$$(\mathbf{R}, \tau) \equiv (\mathbf{R}, \{\{i\}, \{i \cap j\}, \{i \cap j \cap k\}, \dots\}) \equiv (\mathbf{R}, \mathcal{G}) \mapsto E^{\text{PIE-ONIOM}}\quad (8)$$

and hence  $E^{\text{PIE-ONIOM}}$  depends on  $(\mathbf{R}, \tau)$ ; that is,

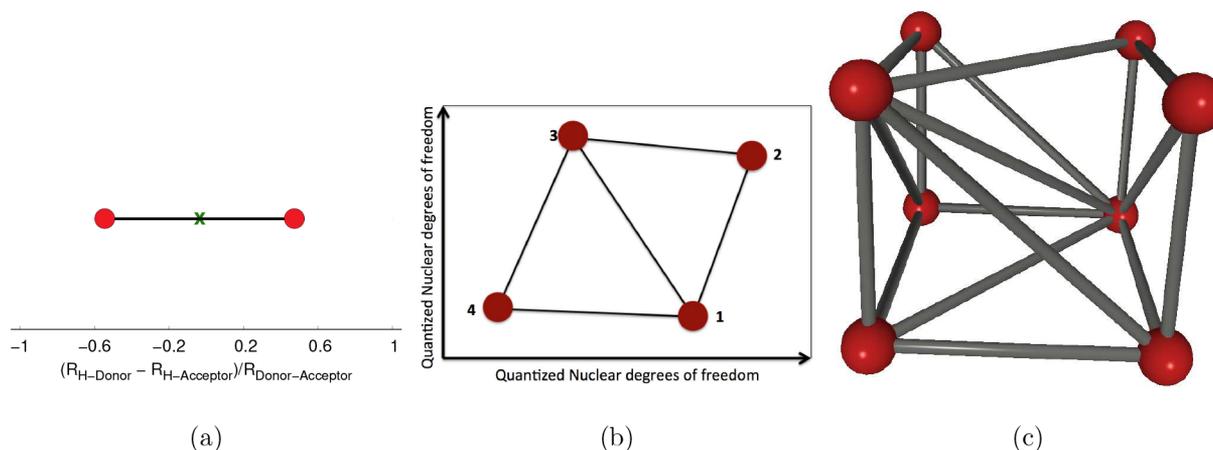
$$E^{\text{PIE-ONIOM}} \equiv E^{\text{PIE-ONIOM}}(\mathbf{R}, \tau)\quad (9)$$

But, it is also possible to define multiple graphs for any given molecular configuration, as is already apparent from the discussion surrounding Figures 1 and 2, and the energy associated with a given graph or topological space definition is

$$(\mathbf{R}, \tau_{\alpha}) \equiv (\mathbf{R}, \mathcal{G}_{\alpha}) \mapsto E_{\alpha}^{\text{PIE-ONIOM}}(\mathbf{R}, \tau_{\alpha})\quad (10)$$

Hence, it becomes necessary to think of the energy of the system as a probabilistic sum over multiple fragmentation topologies (or graphs) since in some sense each of the graphs force a certain kind of locality in the electronic structure and hence may be loosely considered as “valence bond”<sup>76–78</sup> constructs or “diabatic states”.<sup>79–87</sup> Hence,

$$\langle E(\mathbf{R}) \rangle = \sum_{\alpha} \rho_{\alpha}(\mathbf{R}) E_{\alpha}^{\text{PIE-ONIOM}}(\mathbf{R})\quad (11)$$



**Figure 4.** (a) Centroidal decomposition for a one-dimensional potential associated with one proton stepped along the respective donor–acceptor axis in the water–wire studied in this publication. The centroids,  $\{\mathbf{R}_{\text{QM}}^{0,\alpha}\}$  associated with the two diabatic topological states are depicted using solid red circles. The connection between the  $\{\mathbf{R}_{\text{QM}}^{0,\alpha}\}$  is depicted by the edge connecting the two nodes. Similarly, panel b provides an illustration of the centroidal decomposition for a two-dimensional potential associated with two protons stepped along the respective donor–acceptor axes. The centroids,  $\{\mathbf{R}_{\text{QM}}^{0,\alpha}\}$  associated with the four diabatic topological states are depicted using solid red circles. See text for details. This kind of scheme is readily generalizable to higher dimensions as is clear from the Delaunay triangulation of eight centroids in three dimensions in panel c and will be the subject of future publications.

where  $\rho_\alpha(\mathbf{R})$  is the probability corresponding to topology  $\tau_\alpha$  with energy  $E_\alpha^{\text{PIE-ONIOM}}$ , both of which are functions of the nuclear coordinates. Note that topology  $\tau_\alpha$  represents a certain fragmentation protocol and certain connectivity (or adjacency matrix) depicted by the graph  $\mathcal{G}_\alpha$ . A different topology,  $\tau_\beta$ , has a different fragmentation and, hence, different energy and nuclear gradients (because of eq 8). It is further important to clarify that changes in connectivity lead to a change in the adjacency matrix in a graph. This is essentially what happens during dynamics and conformational changes in potential energy surface calculations. Furthermore, it is also possible for the nature of vertices in the graph-theoretic description to change, for example, when a water molecule gains a proton. These subtle changes are depicted in Figure 3 by coloring the nodes. All of the changes are referred to as topological changes, or topological hops, for the remainder of the work. Furthermore, the graphs and topologies,  $\{\mathcal{G}_\alpha\}$  and  $\{\tau_\alpha\}$ , may also be considered as akin to *valence bond states*,<sup>76–78,88–94</sup> and hence the potential energy surfaces constructed by using any individual  $\mathcal{G}_\alpha$  or  $\tau_\alpha$  are interpreted here as *crude diabatic states*.<sup>79–86,95</sup> In the next subsection we introduce a Hamiltonian formalism to obtain the topological weights  $\{\rho_\alpha\}$ , and this process is deeply influenced by the literature on nonadiabatic dynamics.<sup>79,82–84,95–98</sup>

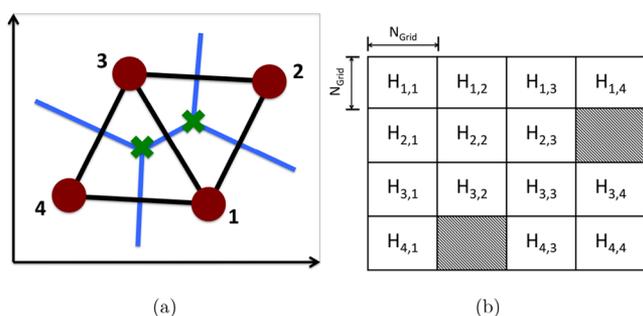
**II.B. Hamiltonian Formalism for Superposition of Graphs,  $\{\mathcal{G}_\alpha\}$  (or  $\{\tau_\alpha\}$ ), To Compute  $\{\rho_\alpha\}$ .** As noted above, we interpret the fragment energy dependence on topologies; i.e.,  $E_\alpha^{\text{PIE-ONIOM}} \equiv E^{\text{PIE-ONIOM}}(\mathbf{R}, \tau_\alpha) \equiv E^{\text{PIE-ONIOM}}(\mathbf{R}, \mathcal{G}_\alpha)$ , as diabatic states,<sup>79–86</sup> and here, for reduced dimensional surfaces, these are treated as functions of some subset of the nuclear degrees of freedom,  $\mathbf{R}_{\text{QM}}$ . In classical dynamics, this nuclear coordinate dependence may be explicitly mapped to the simulation time. However, in quantum nuclear dynamics and potential energy surface calculations, the subspace  $\{\mathbf{R}_{\text{QM}}\}$  may be stepped in some predetermined fashion to obtain surfaces  $\{E_\alpha^{\text{PIE-ONIOM}}(\mathbf{R}_{\text{QM}})\}$  using the map represented in eq 10. It is through this  $\tau_\alpha$  dependence of energy that we interpret these fragmentation-based potential surfaces as crude diabatic

states. Here we develop a general method for interpolation across an arbitrary number of such states, in the sense of eq 11. However, before we develop the interpolation procedure, it is first necessary to describe a *domain decomposition scheme* that defines “local”-connected domains in  $\{\mathbf{R}_{\text{QM}}\}$  that are each optimally described separately by one among a set of graphs,  $\{\mathcal{G}_\alpha\}$ . This domain decomposition scheme is discussed in subsection II.B.1.

While the goal for this publication is to develop smooth potential surfaces from molecular fragmentation, continuous forces for classical dynamics and geometry optimization may also be obtained through straightforward differentiation of eq 11. In such a situation, one would compute Hellman–Feynman forces from eq 11, where the individual diabatic states would have similar gradients as those discussed in ref 2, and the derivative for coefficients,  $\{\rho_\alpha(\mathbf{R})\}$  may be worked out by following the evolution of their time series as done in many semiclassical nonadiabatic dynamics methods.<sup>95</sup> However, we also note that we have conducted classical simulations of larger protonated clusters through direct use of eq 4. In our experience, any discontinuities in energies and forces can be largely mollified by choosing an appropriately large enough value of  $R$  in eq 4 and a large enough spatial range for computing edges for the graph (see the discussion following eq 5.) These aspects regarding classical dynamics will be discussed in future publications.

**II.B.1. Classification (through Unsupervised Learning) of the Nuclear Coordinate Space Grid through Similarities in  $\{\mathcal{G}_\alpha\}$ .** Graph similarity is a well-researched topic of great relevance to machine learning.<sup>99</sup> The general method is to pose the question in terms of similarities in the respective adjacency matrices of two graphs. Here we pursue a simpler approach allowing for improved measures to be incorporated in the future. Details of this method are discussed in Appendix A. The spatial decomposition described in Appendix A is essentially a classification method based on the Forgy method<sup>100</sup> where, at the end,  $\mathcal{N}$  points serve as “centroids”,  $\{\mathbf{R}_{\text{QM}}^{0,\alpha}\}$  of separated domains in  $\{\mathbf{R}_{\text{QM}}\}$ . An example of such a procedure is provided in Figure 4a, where a proton is stepped

along its donor–acceptor axis. The nodes (red circles) represent the centroids for the topologies, and the connection between these nodes indicate the correlation between topologies and contribute to off-diagonal elements in the Hamiltonian discussed in subsection II.B.2. Panels b and c of Figure 4 present the multiple topology centroids (nodes) and their connections (edges) in 2D and 3D potential energy surface domains, respectively. A pictorial representation of the multitopology Hamiltonian (discussed in subsection II.B.2) formed for the 2D case in Figure 4b is represented by Figure 5b.



**Figure 5.** The triangulation procedure (Figure (a)) and the associated, resultant Hamiltonian matrix (Figure (b)). The Hamiltonian that depicts the correlations between the topologies described in Figure 5a (and Figure 4b) is presented in Figure 5b). Blue lines represent the boundaries obtained after the domain decomposition based on  $\{\mathcal{G}_\alpha\}$ . Green crosses are the centroids of triangles containing them. Hamiltonian that depicts the correlations in panel a is presented in panel b.

Before we delve into the details of the Hamiltonian formalism that allows us to interpolate between the various topological descriptions of the electronic structure, we first make some overarching arguments pertaining to the connections between the nodal structure in Figures 4 and 3. Figure 3 shows a hypergraph representation of a single molecular structure. However, as this molecular structure evolves (either in dynamics or during potential surface construction), the graphical network changes and each separate graphical network is represented as a single node in Figure 4. Thus, each node in Figure 4 is representative of one hypergraph such as that in Figure 3; a different node represents a different hypergraph characterized by different connectivities. Thus, each node in Figure 4 pertains to an embedded hypergraph that depicts the electronic structure at that node, and in the Voronoi region (an example shown using blue boundaries in Figure 5a), or  $k$ -means cluster, represented by that node. Thus, the edges, the faces, and the higher order simplexes in Figure 4 depict how the nodes combine within the Hamiltonian formalism (discussed in subsection II.B.2) to yield linear combinations of the electronic structure depicted by each node, to yield a potential surface. As we will see, near the Voronoi separations, the Hamiltonian formalism will allow us to smoothly interpolate, thus morphing uniformly from one topological description to another. As a result the prescriptions below are called continuous topological morphism (abbreviated as CTM).

**II.B.2. Hamiltonian Formalism for Coupled Multitopology Potential Surfaces.** For  $\mathcal{N}$  topologies or graphs given by  $\{\mathcal{G}_\alpha\}$  or  $\{\tau_\alpha\}$ , the fragmentation-based potential energies corresponding to these topologies yield a set of surfaces,

$\{E_\alpha^{\text{PIE-ONIOM}}(\mathbf{R}_{\text{QM}})\}$ . The “multiconfiguration” probabilities,  $\{\rho_\alpha\}$  in eq 11, then allow an interpolation across these states, and to achieve this, we introduce a “Hamiltonian” matrix,  $\mathcal{H}$ , using the following rules:

(1) The Hamiltonian matrix,  $\mathcal{H}$ , is an explicit function of  $\mathbf{R}_{\text{QM}}$ , that is,  $\mathcal{H}(\mathbf{R}_{\text{QM}})$ .

(2) The diagonal elements of  $\mathcal{H}$  at each value of  $\mathbf{R}_{\text{QM}}$  are the topological energies. Hence  $\mathcal{H}_{\alpha,\alpha}(\mathbf{R}_{\text{QM}}) \equiv E_\alpha^{\text{PIE-ONIOM}}(\mathbf{R}_{\text{QM}})$ . For example, as per the domain decomposition procedure discussed above, each grid point  $\mathbf{R}_{\text{QM}}$  is assigned to a centroid  $\mathbf{R}_{\text{QM}}^{0,\alpha}$ . The energy corresponding to the associated topology, as well as energies corresponding to neighboring topologies as defined by the Delaunay triangulation procedure, are the diagonal elements of  $\mathcal{H}$ .

(3) The off-diagonal elements,  $\mathcal{H}_{\alpha,\beta}$ , are computed on the basis of a vicinity parameter,  $\epsilon_{\alpha,\beta}$ . For all sets of surfaces  $\{E_\alpha^{\text{PIE-ONIOM}}(\mathbf{R}_{\text{QM}})\}$  that reside within a vicinity  $\epsilon_{\alpha,\beta}$  from the center of a pair of centroids  $\mathbf{R}_{\text{QM}}^{0,\alpha}$  and  $\mathbf{R}_{\text{QM}}^{0,\beta}$  that is, for

$$\left| \mathbf{R}_{\text{QM}} - \frac{1}{2}(\mathbf{R}_{\text{QM}}^{0,\alpha} + \mathbf{R}_{\text{QM}}^{0,\beta}) \right| < \epsilon_{\alpha,\beta} \quad (12)$$

$\mathcal{H}_{\alpha,\beta}$  is defined as

$$\mathcal{H}_{\alpha,\beta} = \sqrt{\eta(v_{\alpha,\beta})(\eta(v_{\alpha,\beta}) - 1)} (E_\alpha^{\text{PIE-ONIOM}} - E_\beta^{\text{PIE-ONIOM}}) \quad (13)$$

where the function  $\eta(v_{\alpha,\beta})$  is a smooth Heaviside function with parametrization:

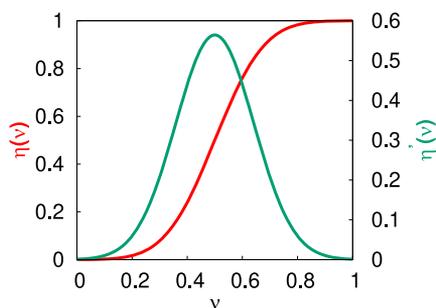
$$\eta(v_{\alpha,\beta}) = c(\text{erf}(b(v_{\alpha,\beta} - a)) + 1) \quad (14)$$

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (15)$$

The parameters  $a$ ,  $b$ , and  $c$  are translation, dilation, and modulation parameters, respectively, as allowed by multiwavelet theories,<sup>101–107</sup> and

$$v_{\alpha,\beta} = \frac{\left| \mathbf{R}_{\text{QM}} - \frac{1}{2}(\mathbf{R}_{\text{QM}}^{0,\alpha} + \mathbf{R}_{\text{QM}}^{0,\beta}) \right| + \epsilon_{\alpha,\beta}}{2\epsilon_{\alpha,\beta}} \quad (16)$$

While we considered many values for parameters  $a$ ,  $b$ , and  $c$  in our studies, including the case where  $b \rightarrow \infty$  and eq 14 tends to a step function, here we present the case when these parameters  $a$ ,  $b$ , and  $c$  are chosen as 0.5, 5, and 0.5, respectively, such that when  $v_{\alpha,\beta}$  is in the window (0,1), then  $\eta(v_{\alpha,\beta}) \in (0, 1)$ . These parameters were chosen as a compromise between smoothness and spread. Figure 6 shows  $\eta(v)$  and its first derivative with respect to  $v$  to illustrate the smoothness and the spread. It should be noted that in eq 12 and eq 16 the seam of intersection between the surfaces obtained using  $\mathcal{G}_\alpha$  and  $\mathcal{G}_\beta$  is approximated to be at the center of the line joining the centroids  $\mathbf{R}_{\text{QM}}^{0,\alpha}$  and  $\mathbf{R}_{\text{QM}}^{0,\beta}$ . Clearly this is an approximation as there is a real seam region that separates the potential surfaces obtained by using  $\mathcal{G}_\alpha$  and  $\mathcal{G}_\beta$ . While, in one dimension, this location may be determined, in multiple dimensions the location and definition of the seam of intersection is not possible to determine and hence approximations are necessary. In section II.B.3 we provide a general description of this problem for arbitrary dimensions, and detailed benchmarks on using such approximate seam regions are discussed in section III.C.



**Figure 6.** (Red curve) Modified error function utilized to construct off-diagonal elements of the Hamiltonian matrix as described in eq 13. (Green curve) First derivative of the function represented by the red curve.

(4) Since  $\eta(v_{\alpha,\beta}) \in (0, 1)$ , all the off-diagonal elements of  $\mathcal{H}$  are purely imaginary. For a set of topological energies  $\{E_{\alpha}^{\text{PIE-ONIOM}}\}$ , the choice of a Hermitian matrix  $\mathcal{H}$  leads to a lowest eigenvalue of  $\mathcal{H}$  outside the span of  $\{\mathcal{H}_{\alpha,\alpha}\}$ . Thus, the interpolated potential energy surface will be lower in energy than all the individual topology surfaces. Instead, it is desired to have the lowest eigenvalue of  $\mathcal{H}$  to reside within the span of  $\{\mathcal{H}_{\alpha,\alpha}\}$ . The purely imaginary off-diagonal element along with a symmetric condition on  $\mathcal{H}$  (that is,  $\mathcal{H}$  is non-Hermitian but special in the sense that it is the sum of a diagonal matrix and an anti-Hermitian matrix) leads to the lowest eigenvalue residing inside the span of all diagonal elements of  $\mathcal{H}$ . In the end, a complex eigenvalue solver is used to calculate all of the eigenvalues, and the lowest eigenvalue is chosen to be the ground state potential energy of the system. In practice, all eigenvalues are found to be real.

The overall structure of such a multitopology Hamiltonian may be understood from the illustration in Figure 5b for the centroidal decomposition in Figure 5a.

**II.B.3. Adaptation of Preceding Ideas to Arbitrary Dimensions.** In subsection II.B.1 and Appendix A we provided a domain tessellation of  $\{\mathbf{R}_{\text{QM}}\}$ , based on the similarities in  $\{\mathcal{G}_{\alpha}\}$  in arbitrary dimensions. Each of the multidimensional-tile centroids,  $\{\mathbf{R}_{\text{QM}}^{0,\alpha}\}$ , is captured uniquely by a single  $E_{\alpha}^{\text{PIE-ONIOM}}$ , and hence we have a family of domains, each one corresponding to a different “diabatic state”. We thus compute  $N$  potential energy surfaces on  $\{\mathbf{R}_{\text{QM}}\}$ , each one corresponding to a distinct topology,  $\tau_{\omega}$  or graph,  $\mathcal{G}_{\alpha}$ . In subsection II.B.2, we consider a Hamiltonian formalism to suitably interpolate between  $E_{\alpha}^{\text{PIE-ONIOM}}$  across the boundary regions of the tiles specified by the centroids  $\{\mathbf{R}_{\text{QM}}^{0,\alpha}\}$ . But we must recognize that there exists a certain degree of arbitrariness in the scale utilized to depict the respective values of  $E_{\alpha}^{\text{PIE-ONIOM}}$ . What is implied here is that since each of the  $E_{\alpha}^{\text{PIE-ONIOM}}$  arises from a different topology,  $\tau_{\omega}$  or graph,  $\mathcal{G}_{\alpha}$  a given grid point on  $\mathbf{R}_{\text{QM}}$  will have different energy values depending on whether a graph  $\mathcal{G}_{\alpha}$  or a graph  $\mathcal{G}_{\beta}$  is used to depict its topology and derive its fragmentation. Thus, there exists an arbitrariness of scale in the family of energies  $\{E_{\alpha}^{\text{PIE-ONIOM}}\}$  and in this section we outline a general multidimensional protocol to overcome this problem.

For a Delaunay-triangulated domain in  $N$ -dimensions (as depicted for 2D and 3D in Figure 4b,c, respectively), we find the set of all of the highest order simplices and obtain their corresponding centroids using the nodes present in them. The potential energy surfaces corresponding to each of the highest order simplices are then shifted to have the same potential

energy at the centroid of their corresponding simplex. For example, for the 2D case shown in Figure 4b, the highest order simplices are triangles,  $\Delta_{123}$  and  $\Delta_{134}$ , respectively. The potential energy surfaces corresponding to nodes “1”, “2”, and “3” will be shifted to have the same potential energy at the centroid of  $\Delta_{123}$ . Independently, the potential energy surfaces corresponding to nodes 1, 3, and 4 will be shifted to have the same potential energy at the centroid of  $\Delta_{134}$ . The centroids  $\Delta_{123}$  and  $\Delta_{134}$  are shown in Figure 5a using green “X”-marks. The Hamiltonian formalism described in subsection II.B.2 is applied to each of these shifted sets of potential surfaces corresponding to  $\Delta_{123}$  and  $\Delta_{134}$ , thus obtaining interpolated surfaces corresponding to “nodes”  $\Delta_{123}$  and  $\Delta_{134}$ . The two interpolated surfaces are then shifted to meet at the center of two centroids, i.e., the midpoint between the two green X-marks in Figure 5a, and the cycle continues until we are left with only one node.

The correlations between the topologies shown in Figure 5a are presented in Figure 5b as a Hamiltonian matrix,  $\mathcal{H}$ . The blocks,  $\mathcal{H}_{ij}$ , represent the coupling between nodes (that is edges in Figure 4b), and the hashed out regions are zero, where for example there are no edge connecting nodes 4 and 2 in Figure 4b and hence the appropriate coupling is hashed out and is zero. Each block is an  $N_{\text{Grid}} \times N_{\text{Grid}}$  matrix, where  $N_{\text{Grid}}$  is the number of grid points.

**II.C. Formal Scaling in Fragment-Based Multitopology Potential Surface Calculations.** We use eq 4 to deduce the scaling in these calculations. First, it is assumed that a molecular assembly has been subdivided into  $N$  monomers (or nodes). Furthermore, the number of basis functions for each node is assumed to be  $M$ , and this affects the computational complexity involved in calculations pertaining to each node. As a result the total number of basis functions in the entire system is roughly the product  $[N \times M]$ .

Equation 4 has multiple segments that are treated in a parallel fashion in our C++ module capable of using multiple electronic structure packages simultaneously, for each fragment-based energy and force calculation. Consequently, the  $E^{\text{level},0}$  calculations in eq 4 scale as

$$\text{Scaling}_{E^{\text{level},0}} = O([MM]^L) \quad (17)$$

where “L” above refers to the polynomial scaling for the lower level of theory, which for DFT is approximately 3.5.

If calculations are performed in a parallel manner, the overall scaling of the second term in eq 4 (nodes, edges, and so on) is bounded by the effort for the maximum rank. Hence, if we assume that the algebraic scaling factor for the higher level of theory is “H” (which for CCSD is 6 and MP2 is 5), the overall scaling for the computation in eq 4 is given by

$$\begin{aligned} \text{Scaling}_{\text{Eq 4}} &= O([MN]^L) + O(\max_{\{\alpha\}} \{\max_{\{r\}} \{(Mr)^H, (Mr)^L\}\}) \\ &= O([MN]^L) + O(\max_{\{\alpha\}} \{\max_{\{r\}} \{(Mr)^H\}\}) \\ &= O([MN]^L) + O(\max_{\{\alpha\}} \{(MR)^H\}) \\ &= O([MN]^L) + O([MR]^H) \end{aligned} \quad (18)$$

It must further be noted that the quantity  $M$  is the number of basis functions per node and is not system size dependent. Thus, the higher scaling second term in eq 18 adds a constant effort and the dominating system size dependent computa-

tional effort arises from the first, lower scaling term in eq 18. As a result,

$$\text{Scaling}_{\text{Eq 4}} \xrightarrow{\text{large } N} \mathcal{O}([MN]^L) \quad (19)$$

As a result, in refs 3–5, we state that eq 4 provides CCSD and MP2 (or post-Hartree–Fock in general) accuracy at DFT cost.

If, on the contrary, the calculations are performed in a serial fashion, then the overall scaling approaches

$$\begin{aligned} \text{Scaling}_{\text{Eq 4}} &= \mathcal{O}([MN]^L) + \sum_{r=0}^R r \sum_{\alpha}^{\text{rank}} \mathcal{O}([Mr]^\alpha) + \mathcal{O}([Mr]^L) \end{aligned} \quad (20)$$

In our calculations, the sum over  $\alpha$  is defined on the basis of a local neighborhood interaction and hence leads to a constant number of terms that is independent of  $N$  and  $R$ . Thus, in the serial case also, the overall scaling tends to that in eq 19.

**II.C.1. Formal Scaling for Continuous Topological Morphism.** The situation for CTM is more complex in the sense that the computation in eq 4 is to be performed at multiple grid positions ( $N_{\text{Grid}}$  in number), and for multiple topologies,  $N_{\text{Top}}$  in number. Since the sampling functions discussed in refs 108–112 would reduce the value of  $N_{\text{Grid}}$  by several orders of magnitude, we assume the new set of grid points to be  $N_{\text{Grid}}^{\text{Shannon}} \ll N_{\text{Grid}}$ . However, for each additional grid point, the computational effort includes, calculation of  $E^{\text{level},0}$  (that is,  $\mathcal{O}([MN]^L)$ ) and only a subset of the rank- $R$  fragments that contain dimensions that are perturbed along the grid. Thus, this latter aspect is expected to have critical impact in computing reduced dimensional surfaces. Thus, computing a single potential surface approaches the complexity:

$$\text{Scaling}_{\text{Eq 4-OnePEs}} \xrightarrow{\text{large } N} \mathcal{O}(N_{\text{Grid}}^{\text{Shannon}} \times [M \times N]^L) \quad (21)$$

where, in principle, the actual scaling is only determined by the second L-scaling part, but in practice  $N_{\text{Grid}}^{\text{Shannon}}$  can be a large number.

Now, the question of multiple potential surfaces as dictated by the term,  $N_{\text{Top}}$ , and to analyze this we revert back to eq 18. The key idea is that, for every additional topology,  $E^{\text{level},0}$  does not need to be recomputed and can be reused from a previous topology calculation. But, some of the rank- $R$  simplex fragments do need to be computed, and hence the overall complexity of potential surface calculations scales as

$$\text{Scaling}_{\text{CTM}} = N_{\text{Grid}}^{\text{Shannon}} \times \{\mathcal{O}([MN]^L) + N_{\text{Top}} \times \mathcal{O}([MR]^\alpha)\} \quad (22)$$

It is true that molecular system size  $N$  only appears in the first, lower scaling term. It is further true that one may expect the number of topologies to be far fewer than the number of sample points on the potential surface grid; that is,  $N_{\text{Grid}}^{\text{Shannon}} \gg N_{\text{Top}}$ . Hence, in the large system limit, one can make the argument that

$$\text{Scaling}_{\text{CTM}} \xrightarrow{\text{large } N} N_{\text{Grid}}^{\text{Shannon}} \times \{\mathcal{O}([MN]^L)\} \quad (23)$$

but for the system size considered here, and for all current practical purposes, it is safer to assume that the scaling is

governed by eq 22, which states that while a significant portion of the computational effort is dominated by the lower L-scaling (that is DFT scaling), for medium sized systems, there still exists a portion of the calculation that scales as the higher level electronic structure method, with algebraic scaling  $H$ , and this second part is linearly attenuated by the number of topologies required for potential surface generation. *Furthermore, this second higher scaling H-dependent term does not depend on system size and is only controlled by  $M$ , the size of the monomer, and  $R$ , the size of the many-body envelope.* In all cases it is expected that  $N_{\text{Grid}}^{\text{Shannon}} \gg N_{\text{Top}}$ . In practice, the two parts of the computation in eq 22 are done in parallel, and even for the system size treated here, the lower level of theory dominates the computational effort. This scaling is expected to become more favorable with system size (from an electronic structure perspective) as is clear from eq 22. The large  $N_{\text{Grid}}^{\text{Shannon}}$  prefactor in eq 23 is a critical limitation in quantum nuclear treatment, and this work does not address these concerns. Please see ref 113, where the potential for tensor-network decomposition as a step to further reduce the  $N_{\text{Grid}}^{\text{Shannon}}$  complexity is discussed. Future publications will further expand on these computational gains to present a tractable overall software solution.

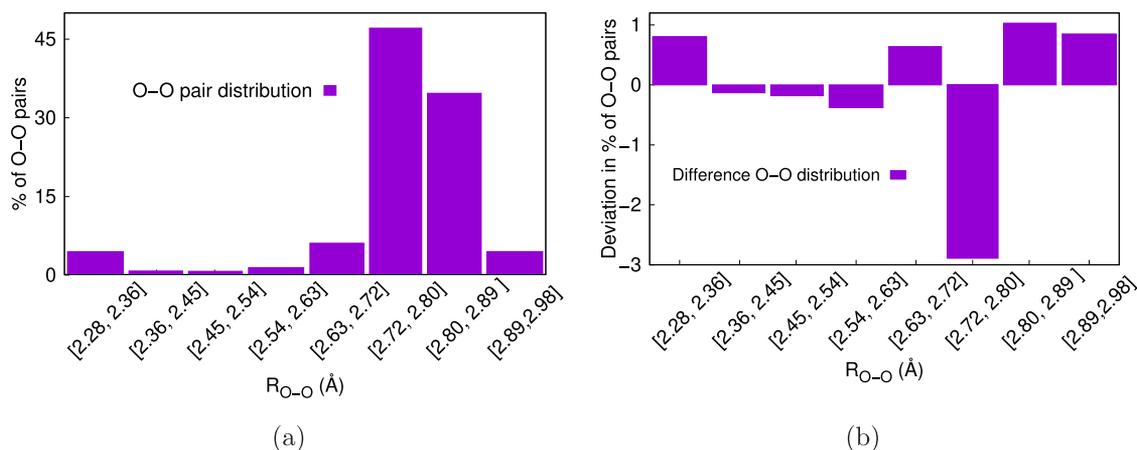
**II.C.2. Computational Aspects.** There are several components to the computational implementation.

(a) The definition of nodes represents a level of coarse-graining and constitutes the first step in the algorithm.

(b) Following this, edges are constructed. Here we enforce two conditions. While these conditions are discussed here in a manner specific to water, the algorithms are general as may be clear from the discussion in ref 4. In our case here, the oxygen atoms for each water represent a node. We then create edges between neighboring nodes based on a shortest distance criterion. Additional edges are added between nodes that are within 10% from the respective shortest distance edge. This is the same protocol as that used in ref 1.

(c) To obtain all fragments from the associated graph, there are three algorithms available for use in our computer program. Two of these algorithms are discussed in detail in previous publications.<sup>1,4</sup> Both of these algorithms are essentially bit-manipulation algorithms, where each fragment obtained from the edges above are treated as bits in an integer. Integer AND operations yield nodes, and electronic structure input files are created in an efficient manner using these bitwise-AND operations. (See Chart 1 of Appendix A in ref 1). A more sophisticated Python script that obtains all objects up to rank- $R$  for a given graph defined by nodes and edges is currently under development and will be discussed in a future publication.

(d) Once the fragments are generated on the basis of the algorithms discussed above, there are two additional simplifications that arise here (i) within the context of calculation of a diabatic state at all grid points and (ii) with computation of multiple topologies at each grid point. In both cases, the common fragments are eliminated and computation is done only for those fragments where the atomic positions change. In this manner the code is optimized to compute only those fragments that are needed at additional grid points and/or additional topologies. Currently this is done through a shell script, but future versions of the code will use graph utilities from Python, steps for which have already been taken as discussed under point c above.



**Figure 7.** (a) Pair oxygen–oxygen distribution function for all geometries in the AIMD trajectory. (b) Deviations for the same distribution for the sample set chosen for reduced dimensional potential energy surface calculations. Panel b shows that the sample set chosen for potential surface calculations faithfully represents the distribution in panel a.

(e) All electronic structure calculations required to assemble the fragmentation energies in eq 4 are performed through an MPI parallelized C++ module that is capable of simultaneously using multiple electronic structure packages within a single  $E_{\text{graph-theoretic}}$  calculation (eq 4). In this work, all components of the right side of eq 4 are computed using the Gaussian series of electronic structure programs.<sup>114</sup>

### III. NUMERICAL BENCHMARKS ON THE ACCURACY OF POTENTIAL SURFACES AND PROTON EIGENSTATES OBTAINED USING MULTIPLE FRAGMENTATION PROTOCOLS

Here we probe the accuracy in computing molecular potential energy surfaces, proton eigenstates, and eigenenergies in agreement with post-Hartree–Fock theories using the  $(\text{H}_2\text{O})_{12}\text{H}^+$  water–wire system shown in Figure 2 as a benchmark. This particular system has been widely studied by several groups using empirical potential models.<sup>54–56,115–117</sup> These protonated water–wires are seen in many constrained environments such as biological membranes and enzyme active sites,<sup>118–120</sup> ion channels,<sup>121</sup> carbon nanotubes,<sup>122–124</sup> and fuel cells.<sup>125</sup> Water–wires are also present in the photosynthetic reaction center of *Rhodobacter sphaeroides*,<sup>119</sup> where they are responsible for proton transfer to a secondary quinone group.<sup>126</sup>

The method outlined above is general and adaptive and requires (a) location of contiguous regions of the potential surface governed through treatment by a single fragmentation protocol or topology,  $\tau_w$  or graph,  $\mathcal{G}_w$ , (b) the associated set of centroids,  $\mathbf{R}_{\text{QM}}^{0,\alpha}$  that represent these contiguous regions, and (c) the intervening intersections or seam regions that separate these multiple topological regions. While these intervening seam regions may be known as part of classical dynamics simulations and one-dimensional reaction path calculations, since changes in topologies would be accompanied by spikes in electronic energy, such information may not be readily available when multidimensional potential surfaces are computed. Furthermore, the regions defining the intersection seam may themselves be multidimensional subspaces and thus inherently complex to determine. In subsection II.B.3, we have provided a general, multidimensional protocol for computing the interpolated surfaces without knowledge of the seam. However, one critical question that arises in the algorithm is

how crucial is the precise identification of the seam region for effective interpolation across multiple topological surfaces, and in subsection III.C, we present benchmarks that indicate that this information is not necessary to be known in an extremely precise fashion. In fact, our tests here indicate that errors in seam detection of the order of 0.3 Å in one-dimensional studies yield a negligible change in potential surfaces.

In subsection III.A, we present computational aspects relating to fragment and graph (or topology) definitions and other parameters used for multitopology interpolation. Accuracy of potential surfaces and associated eigenstates and eigenenergies computed using multitopology interpolation where the seam regions are known are discussed in subsection III.B and, when they are not known, are discussed in subsection III.C; in subsection III.C we also present benchmarks for the general approach in subsection II.B.3. The latter is especially critical, as noted above, for multidimensional cases. Multidimensional aspects are briefly discussed in subsection III.D with more details to be considered in future publications.

**III.A. Water–Wire Geometries for Reduced Dimensional Quantum Nuclear Potential Surfaces, Fragmentation Protocols, and Other Computational Details.** We gather a realistic range of shared hydrogen potential surfaces for benchmark calculations, by first examining results from ab initio molecular dynamics trajectories. A set of ten different molecular geometries are assembled from AIMD studies on  $(\text{H}_2\text{O})_{12}\text{H}^+$  performed at the B3LYP/6-31++g(d,p) level of electronic structure theory; these geometries are chosen at roughly 5 fs increments from the AIMD trajectory. The configurational sampling of these structures may be gauged from the oxygen–oxygen pair separation since this dictates the kind of potential energy surface that the shared proton experiences. The oxygen–oxygen pair distribution function obtained from the AIMD trajectory is shown in Figure 7a, whereas the associated pair distributions chosen for quantum nuclear studies deviates from Figure 7a as given in Figure 7b. The relevance of the distribution in Figure 7a is critical to note. As seen from several previous studies,<sup>127–129</sup> the short oxygen–oxygen distances (of the order of 2.5 Å or less) correspond to short strong, Zundel-like<sup>130–134</sup> hydrogen bonds.<sup>135,136</sup> On the contrary, oxygen–oxygen distances in the range of 2.8 Å are seen in more delocalized Eigen-like

systems,<sup>137,138</sup> in larger water clusters<sup>127</sup> and in condensed phase studies.<sup>59,139</sup>

In the protonated water–wire (H<sub>2</sub>O)<sub>12</sub>H<sup>+</sup> examined here, there are 11 nearest neighboring O–O pairs all containing a shared proton. We compute one-dimensional potential energy surfaces for these 11 protons along reduced dimensions directed along the donor–acceptor oxygen pairs using the multiple fragmentation topology based electronic structure method outlined above. From our sampled data set of 110 potentials (10 geometries with 11 potential surfaces from each geometry), we find that any potential energy surface belonging in the O–O distance range of 2.28–2.72 Å has approximately a 93% probability of containing a topological hop, due to the short–strong nature of the associated hydrogen bond, whereas a geometry belonging in the range of 2.72–2.80 Å has approximately a 25% probability of containing topological hops. Oxygen–oxygen separations in the range 2.80–2.98 Å generally are well-characterized by a single fragmentation protocol, and the multitopology formalism is not necessary in these cases. Thus, the O–O pair distribution not only classifies the oxygen pairs but also the corresponding potential energy surfaces computed in this publication. Hence our analysis follows the classification provided by Figure 7.

Fragmentation of all of the protonated water clusters is performed in such a way that primary fragments (or edges in the graphical representation) are chosen to be water dimers or Zundel cations, H<sub>5</sub>O<sub>2</sub><sup>+</sup>, and their overlaps (or nodes in the graphical representation) are either hydronium ions (H<sub>3</sub>O<sup>+</sup>) or water molecules. See Figure 3b. Parameters employed during the potential surface calculations are shown in Table 1. The

**Table 1. Parameters Used To Construct Fragment-Based Potential Surfaces**

characteristic	value
no. of grid points in each dimension	149
spatial spread of the grid along donor–acceptor direction <sup>a</sup>	0.8 (or 1.0) Å
grid spacing along donor–acceptor direction	0.005 (or 0.007) Å
lower level of theory	B3LYP/6-31++g(d,p)
target higher level of theory	MP2/6-31++g(d,p)
topological vicinity parameter, $2\epsilon_{\alpha,\beta}$ in eq 12	0.4 Å

<sup>a</sup>Grid spreads are chosen on the basis of the the donor–acceptor distances.

computations are performed using an MPI parallelized C++ driver that computes the energy and forces, by spawning fragment calculations that are processed using the Gaussian series of electronic structure programs<sup>114</sup> for electronic energy and gradient calculations at independent compute nodes. But first the centroids are computed for each fragmentation topology. An illustration of the associated Delaunay triangulation<sup>65,69,70</sup> and Voronoi diagram<sup>66–68</sup> of the potential surface constructed using the centroids for each topology as vertices for the Delaunay triangulation is shown in Figure 4. Next, energy surfaces associated with each topology are constructed, and here the use of sampling functions<sup>108–111</sup> will have the great impact in reducing the number of electronic structure calculations. The multitopology interpolated surfaces are constructed using the algorithms explained above. Benchmarks are provided in the following sections.

### III.B. Interpolation across Surfaces Obtained from Multiple Topologies When the Intersection Seam Is

**Already Known.** The accuracy of approximate potential surfaces computed from the above formalism is gauged by comparison with a reference calculation performed for the full system at the target higher level of theory which in this case is chosen to be MP2/6-31++g(d,p) (see Table 1). Error estimates for the interpolated potentials are computed according to the relation

$$\epsilon_{V,w_j} = \sum_i w_j(r_i) |V^{\text{app}}(r_i) - V^{\text{high}}(r_i)| \quad (24)$$

where  $V^{\text{app}}(r_i)$  is the approximate potential computed from multiple fragment topologies and  $V^{\text{high}}$  signifies the same at the higher (target) level of electronic structure theory. Three normalized weighting functions are employed where each one selectively gauges the error at different regions of the potential. The ground eigenstate weighted mean absolute error is obtained according to the weighting function

$$w_1(r_i) = |\psi_0(r_i)|^2 \quad (25)$$

where  $\psi_0$  is the ground eigenstate and gauges the error between  $V^{\text{app}}(r)$  and  $V^{\text{high}}(r)$  in critical regions of the potential populated by the ground eigenstate. A Heaviside function may be used to gauge the error in a given energy window, and the associated weighting function for energies less than 15 kcal/mol may be written as

$$w_2(r_i) \propto \begin{cases} 1, & \text{if } [V^{\text{high}}(r_i) - \min_{V_i} V^{\text{high}}(r_i)] < 15 \text{ kcal/mol} \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

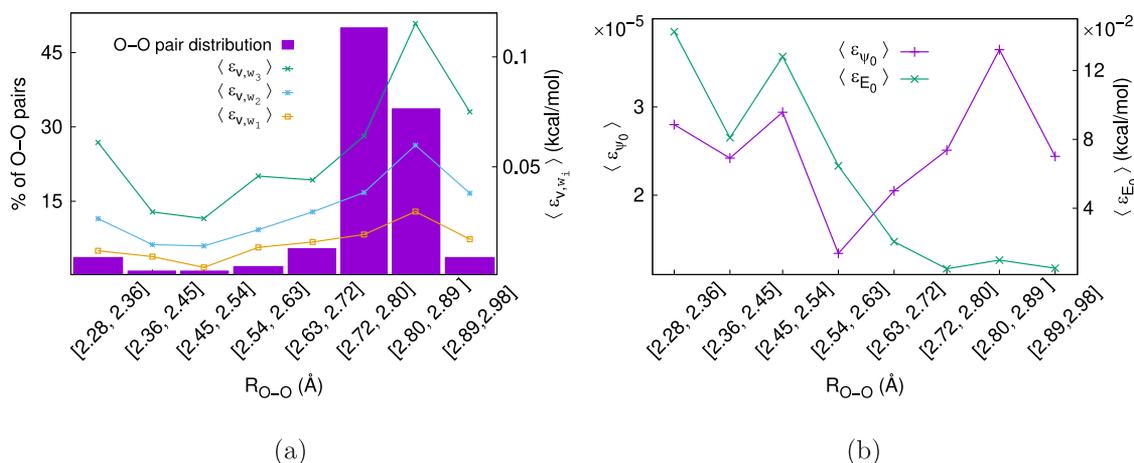
Furthermore, an estimate of the global error in  $V^{\text{app}}(r)$  may be computed according to

$$w_3(r_i) = \frac{1}{N} \quad (27)$$

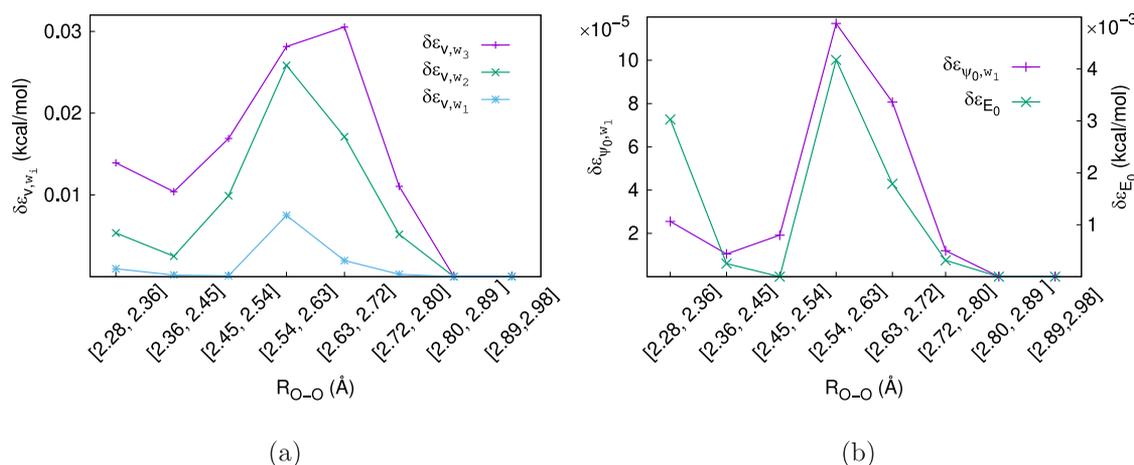
and in all cases the weighting functions are L<sup>1</sup>-normalized to unity and hence  $N$  in eq 27 is the number of grid points. Furthermore, these errors are hydrogen bond donor–acceptor distance,  $R_{\text{DA}}$ , dependent, and hence we average these  $\epsilon_{V,w_j}$  values in our studies below over the distribution of donor–acceptor distances inside each bin shown in Figure 7, and these averages are represented as  $\langle \epsilon_{V,w_j} \rangle$  and are functions of the oxygen–oxygen donor–acceptor distances. In Figure 8a, we present the O–O pair distribution on the left vertical axis and the average errors in potential,  $\langle \epsilon_{V,w_j} \rangle$ , on the right vertical axis. Thus, any point on the  $\langle \epsilon_{V,w_1} \rangle$  curve plotted on the right vertical axis represents the average error over all  $\epsilon_{V,w_1}$  errors for the potential surfaces belonging to a specific donor–acceptor bin represented along the horizontal axis. The bin averaged potential surface errors over the 110 reduced dimensional potential surfaces used here are well within 0.15 kcal/mol. Specifically, the ground eigenstate weighted error, which provides the error in the critical region of the potential is well within 0.02 kcal/mol. In all cases, as is to be expected,  $\langle \epsilon_{V,w_1} \rangle < \langle \epsilon_{V,w_2} \rangle < \langle \epsilon_{V,w_3} \rangle$ .

We next examine the errors in eigenvalues and eigenvectors obtained from the potential. Thus, the L<sup>1</sup> norm of errors in the proton ground eigenstate is defined according to

$$\epsilon_{\psi_0} = \frac{1}{N} \int d\vec{r} \|\psi_0^{\text{app}}(\vec{r}) - \psi_0^{\text{high}}(\vec{r})\| \quad (28)$$



**Figure 8.** In Figure 8(a), the histogram depicts the significance of a given O–O distance as is shown on the left vertical axis and as in Figure 7(a). Average potential energy surface errors on the right vertical axis of Figure 8(a) are computed for all potentials in a given bin as discussed in text. Figure 8(b) shows the mean absolute error in proton ground eigenstates for potential surfaces belonging to each bin along the donor–acceptor axis (left axis) and the corresponding errors in the averaged zero-point energies are shown on the right axis.



**Figure 9.** (a) Average of standard deviations of potential surface errors,  $\epsilon_{V,w_i}(\mathbf{R}_{QM}^{Seam})$  when a set of 10 different seam inputs within a grid span of 0.3 Å are used for interpolation. (b) Average of standard deviations for the corresponding ground eigenstates errors,  $\epsilon_{\psi_0}(\mathbf{R}_{QM}^{Seam})$  and ground eigenenergies,  $\epsilon_{E_0}(\mathbf{R}_{QM}^{Seam})$ , for each donor–acceptor bin.

where  $\psi_0^{app}$  and  $\psi_0^{high}$  are proton ground eigenstates obtained using multitopology interpolation and full surface high-level calculation, respectively. The absolute error in the ground eigenenergies are

$$\epsilon_{E_0} = |E_0^{app} - E_0^{high}| \quad (29)$$

where again  $E_0^{app}$  and  $E_0^{high}$  are defined as previously done for the eigenstates. The corresponding donor–acceptor bin averaged values are depicted as  $\langle \epsilon_{\psi_0} \rangle$  and  $\langle \epsilon_{E_0} \rangle$  and again, as in the case of  $\langle \epsilon_{V,w_i} \rangle$ , depend on the oxygen–oxygen hydrogen-bond donor–acceptor distances. Figure 8b shows the average of such errors for all donor–acceptor bins. All errors are well within acceptable range. Thus, the multiple fragmentation-based interpolation schemes show good accuracy at much reduced computational cost.

Electronic structure calculations are performed using MPI parallelization of fragment calculations. For a given topology and a grid  $\{\mathbf{R}_{QM}\}$ , we reduce the topology to only contain the fragments which change their geometries across  $\{\mathbf{R}_{QM}\}$ . For multiple topology calculations, we reuse the common fragment energies across different topologies. Overall, we compute the

fragment-based potential energy surfaces using only 10% of the computational cost as compared to the reference calculation performed for the full system at the target higher level of theory. These computational developments yield further reduction in computational effort in addition to that already available due to the fragmentation protocol. However, as noted previously, all of these calculations were for cases where the seam of the intersection between the topologies was known, and in the next section we address the problem when the seam is unknown as this is intrinsic to the study of multidimensional problems.

**III.C. Interpolation for Cases Where the Seam of the Intersection across Topological States Is Not Known or Not Possible To Calculate.** To probe the effect of knowledge of a seam in the previous section, we first modify eq 16 to read

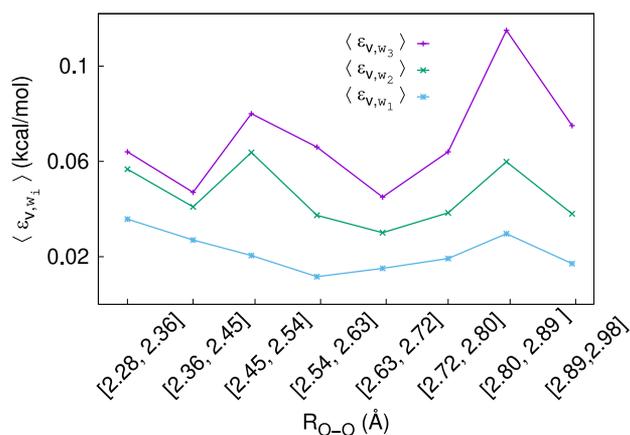
$$v_{\alpha,\beta} = \frac{\mathbf{R}_{QM} - \mathbf{R}_{QM}^{Seam} + \epsilon_{\alpha,\beta}}{2\epsilon_{\alpha,\beta}} \quad (30)$$

where the value of  $\mathbf{R}_{QM}^{Seam}$  is now to be scanned, and with that the accuracy of the interpolation is determined. Toward this,

we choose a range of values for  $\mathbf{R}_{\text{QM}}^{\text{Seam}}$  in a neighborhood of size 0.3 Å for all of the calculations shown in the previous section. The bin averages of standard deviation for  $\epsilon_{V,wj}(\mathbf{R}_{\text{QM}}^{\text{Seam}})$ ,  $\epsilon_{\psi 0}(\mathbf{R}_{\text{QM}}^{\text{Seam}})$ , and  $\epsilon_{E0}(\mathbf{R}_{\text{QM}}^{\text{Seam}})$ , where we have parenthetically noted the dependence on  $\mathbf{R}_{\text{QM}}^{\text{Seam}}$ , is then presented in Figure 9. For potential surfaces, standard deviations are computed for the three types of potential energy surface errors  $\epsilon_{V,w1}$ ,  $\epsilon_{V,w2}$ , and  $\epsilon_{V,w3}$  described earlier, while performing interpolation with different seam inputs within 0.3 Å. Specifically, in Figure 9a we present the average of standard deviations in potential energy surface errors,  $\epsilon_{V,wj}(\mathbf{R}_{\text{QM}}^{\text{Seam}})$ , corresponding to O–O pairs inside each bin. This gives us insight into the effect of not using an exact seam value on the errors obtained for the potential surfaces computed within the range of donor–acceptor distances. For example, a point on a  $\delta\epsilon_{V,w1}$  curve shown in Figure 9a corresponds to the average of standard deviations in  $\epsilon_{V,w1}$  obtained when 10 different seam locations within a grid spread of 0.3 Å were used in the multitopology interpolation calculations. We see that the standard deviations in potential energy surface errors are less than 0.03 kcal/mol in all of the donor–acceptor distances. Thus, the exact location of the seam does not seem to play a significant role in the interpolation. An approximate guess for the seam can be utilized as input to the CTM program to get the interpolated surface, as may be gauged from eq 12.

Similarly, the bin averages for standard deviations in proton eigenstate ( $\delta\epsilon_{\psi 0,w1}$ ) and eigenenergy ( $\delta\epsilon_{E0}$ ) errors are depicted in Figure 9b. It is clear that the averages of the standard deviations in all of the bins are negligible and of the order of  $10^{-5}$  for ground eigenstates and less than 0.01 kcal/mol for the ground state eigenenergy.

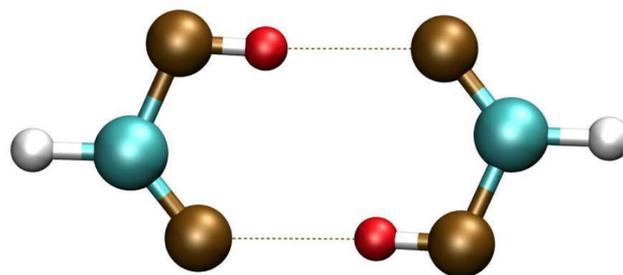
The centroidal Voronoi decomposition of  $\{\mathbf{R}_{\text{QM}}\}$  discussed in subsection II.B.1 suggests the middle point of the topology centroids as an appropriate choice for a seam. We confirm our hypothesis by utilizing the center of topology centroids as the seam input for all of the multitopology CTM interpolated potentials computed in this work and thus computing the bin averaged errors shown in Figure 10 using the same set of geometries as mentioned earlier in Figure 8a. We find that the



**Figure 10.** The average potential energy surface errors computed using the center of topology centroids as seam. The errors are computed in the same way as explained in Figure 8a. Hence, for each of the three curves every point represents the average of errors in the same bin for the CTM interpolated potential energy surfaces in comparison to the reference potential surface computed using a higher level of theory.

errors thus obtained not only follow a similar trend but are also of very similar magnitudes. The bin averaged errors for all of the bins are less than 0.12 kcal/mol, which is the case in the known seam case presented in Figure 8a. Thus, we conclude that circumventing the problem of finding an exact seam via the use of the middle point of the topology centroids serves as a good approximation in CTM. This approximation in locating a seam is specifically crucial for the multidimensional potential surfaces where  $k$ -means provides an efficient algorithm to obtain the centroids,  $\{\mathbf{R}_{\text{QM}}^{0,\alpha}\}$  based upon  $\{\mathcal{G}_\alpha\}$  on a grid  $\{\mathbf{R}_{\text{QM}}\}$  in arbitrary dimensions.

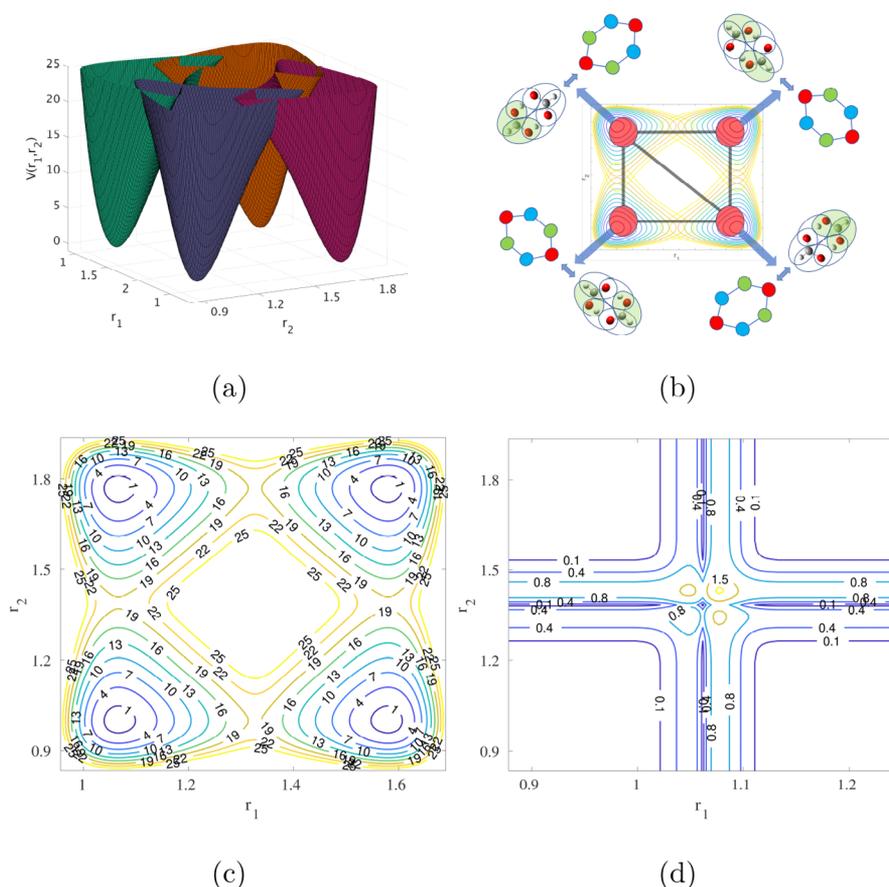
**III.D. Coupled Proton Motion in Cyclic Formic Acid Dimer,  $(\text{HCOOH})_2$ .** Coupled proton dynamics is critical in many chemical and biological processes. These include hydrogen transfer<sup>61,140</sup> reactions in biological<sup>141,142</sup> and atmospheric<sup>143</sup> problems involving secondary isotope effects,<sup>142,144–146</sup> LASER dyes,<sup>147</sup> and several solar energy<sup>148</sup> and fuel cell<sup>149,150</sup> applications. Double proton transfers in the carboxylic acid dimers<sup>151–155</sup> such as the formic acid dimer (FAD; shown in Figure 11) are prototypical exam-



**Figure 11.** Coupled proton transfer for the formic acid dimer.

ples.<sup>151,153,155–157</sup> The FAD has especially been widely studied to quantify the modulation of barrier heights due to the coupling between nuclear degrees of freedom and to computed associated tunneling splittings.<sup>158,159</sup>

In our study we consider the two-dimensional coupled proton stretch degrees of freedom depicted in Figure 11. The true problem is at least three-dimensional and must also include the backbone fluctuations which are ignored in this study, given our goal to demonstrate the multitopology method. Furthermore, given the complexity involved in computing the potential surface across a multiple dimensional grid, here we approximate each proton stretch as a Morse oscillator. The Morse potential energy functions thus utilized are parametrized using ab initio electronic structure calculations performed at MP2/6-31++g(d,p) level of theory. (We have also tested these for a CCSD/6-31g parametrization, where accuracies are obtained similar to those for the MP2 case shown here.) Specifically, the well depth is calculated using the energy difference between the optimized structure and saddle point structure. The optimized structure is the minimum energy structure shown in Figure 11, whereas the saddle point configuration has both hydrogen atoms located at the center of the respective donor–acceptor moieties. The well width parameter utilized in the Morse potential energy function is calculated using the harmonic force constants obtained at the optimized geometry pertaining to the vibrational modes corresponding to the symmetric and asymmetric O–H stretch along the directions shown in Figure 11. The resultant Morse potentials are illustrated in Figure 12a.



**Figure 12.** (a) Four individual diabatic curves; (b) associated graphical representation. Each node (pink spheres, panel b) represents an embedded graph, and this is depicted by expanding on the node to provide both graphical and set-theoretical representations of the molecular framework. (c) Interpolated potential energy surface obtained from multiple topologies; (d) difference between the interpolated potential and the individual diabatic states.

In the language of molecular fragmentation used above these individual Morse oscillator surfaces appear as four different diabatic states (Figure 12a), arising from four different bonding topologies,  $\{\tau_\alpha\}$ , (or graphs,  $\{\mathcal{G}_\alpha\}$ ). The process of converting these diabatic states to a graphical representation, using the methods discussed earlier, is outlined in Figure 12b. Specifically, the individual diabatic states are represented as pink-colored nodes in Figure 12b. These nodes themselves represent graphs that are then displayed in the four corners of Figure 12b. Both set-theoretic and graphical representations are provided. Each set in the representation depicts an edge, whereas the overlap between sets is a node in the graphical decomposition provided beside the set-theoretic fragmentation. Thus, the sets contain either HCO or HCOH fragments that are colored differently. The coloring scheme for the graphs is similarly chosen to maintain consistency. As can be seen, the topological differences arise from the constituent molecular fragments that are present in each set, node, or edge.

We interpolate between the diabatic potentials using our multitopology interpolation scheme explained in subsection II.B. The interpolated curve is shown in Figure 12c, and the difference between the interpolated curve and the individual diabatic curves is presented in Figure 12d. As may be clear from Figure 12d, the interpolated surface smoothly transitions between the various diabatic Morse curves and only differs from each at the intervening seam regions. As seen from Figure 12c, the central part of the contours is a high energy region.

Hence the behavior shown in the central part of Figure 12d is of minimal significance to quantum nuclear effects for this particular arrangement of the formic acid dimer system. Further analysis of the associated eigenvectors and eigenenergies for the multidimensional potentials will be a part of future publications.

#### IV. CONCLUSION

In recent years, molecular fragmentation methods have greatly revolutionized quantum chemistry through the promise of accurate electronic structure at much reduced computational cost accompanied by lower algebraic scaling. There have been several studies constructed relating to accurate relative isomer stabilization studies in systems of biological interest, vibrational frequency calculations, and also ab initio molecular dynamics simulations, to name a few, but methods that can accurately and rigorously account for the dynamical nature of molecular fragmentation, where the molecular fragmentation protocol is to continuously adapt as a function of nuclear framework, have been particularly elusive. This critical gap has limited the progress and application of molecular fragmentation in critical areas such as computing accurate potential energy surfaces<sup>87,160</sup> and reaction pathways,<sup>161</sup> and in making this powerful method available for studies pertaining to quantum nuclear effects. In this work, we have taken a significant step toward precisely these goals. We develop a new adaptive approach to molecular fragmentation where potential energy surfaces can

be accurately computed by simultaneously considering multiple fragmentation protocols.

Starting from a graph-theoretic approach to molecular fragmentation, we further tessellate the molecular potential energy space into contiguous regions each defined by a separate fragmentation protocol. The creation of these contiguous regions in the potential energy space relies on an unsupervised-learning protocol known as the *k*-means; here, a similarity index is computed between the graphs representing different potential energy surface segments and the surface is divided (or classified) on the basis of such similarity. Following this, a Hamiltonian formalism is presented which allows for smooth interpolation across multiple fragmentation protocols by constructing the potential as a superposition of molecular fragmentation topologies. The approach presented is multi-dimensional in nature, and in this case, the physical picture that appears is that of a hypergraph, where each node in the hypergraph codes in the electronic structure in a local region of the potential surface; edges in this hypergraph provide a conduit to correlate and combine these local fragmentation topologies to create the overall surface.

The approach is benchmarked by computing reduced dimensional shared proton potential surfaces in a protonated water–wire system. The surfaces are found to be in very good agreement with post-Hartree–Fock methods such as MP2 (agreement in the sub-kcal/mol range); furthermore, shared proton eigenstates and eigenvalues are also found to be in equally good agreement, and of course, as always in molecular fragmentation, the computational effort is much reduced. Future studies will combine these fragmentation approaches with potential surface sampling efforts<sup>108</sup> to further reduce computational costs and make state of the art, on-the-fly, quantum nuclear dynamics possible.

## ■ APPENDIX A

### Domain Decomposition of the Potential Surface Using Unsupervised Learning

A reduced representation is first constructed for each graph. This is done by assigning an integer value to each graph:  $\mathcal{G}_\alpha \mapsto N_\alpha$ . There may be several ways to do this, and our procedure is as follows: The graph or topology that is adaptively generated for a given nuclear configuration  $\mathbf{R}_{QM}$  is stored as a matrix of bits where the presence of each atom (rows of the matrix) in each edge fragment (columns of the matrix) is indicated by an “ON”-bit (or 1) and the absence is indicated by an “OFF”-bit (or 0). We then obtain a unique integer representation from each of these matrices of bits as  $\left\{ \sum_j 2^{j-1} \left[ \text{sort}_j \sum_i 2^{i-1} a_{ij} \right] \right\}$ , where  $a_{ij}$  is ON when the *i*th atom is present in the *j*th fragment. This transformation leads to a piece-wise constant representation that encodes how  $\mathcal{G}_\alpha$  changes in the subspace  $\{\mathbf{R}_{QM}\}$ . We then utilize an unsupervised machine-learning algorithm called *k*-means,<sup>100</sup> which is similar to the centroidal Voronoi tessellation<sup>162–164</sup> procedure, on the piece-wise constant function to split  $\{\mathbf{R}_{QM}\}$  into  $\mathcal{N}$  (number of unique topologies) distinct connected spatial regions each of which contain only one topology. The spatial decomposition described above is essentially a classification method based on the Forgy method<sup>100</sup> where (a)  $\mathcal{N}$  random points serve as the first estimates for the “centroids”,  $\{\mathbf{R}_{QM}^{0,\alpha}\}$  of the separated domains. (b) Following this, each point in  $\{\mathbf{R}_{QM}\}$  is sequentially assigned to the region

represented by its nearest centroid, which essentially amounts to the construction of a Voronoi diagram<sup>162–164</sup> associated with nodes  $\{\mathbf{R}_{QM}^{0,\alpha}\}$ . (c) The centroid of the resultant tessellation then replaces  $\{\mathbf{R}_{QM}^{0,\alpha}\}$ . The iteration continues until there is no change in the domain decomposition (that is, the  $\{\mathbf{R}_{QM}^{0,\alpha}\}$  converges) and is expected to converge faster for cases containing a large number of points  $\{\mathbf{R}_{QM}^{0,\alpha}\}$ . A Delaunay triangulation<sup>65–68</sup> is then constructed from the family  $\{\mathbf{R}_{QM}^{0,\alpha}\}$ .

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: iyengar@indiana.edu.

### ORCID

Srinivasan S. Iyengar: 0000-0001-6526-2907

### Funding

This research is supported by the National Science Foundation Grant NSF CHE-1665336 to SSI.

### Notes

The authors declare no competing financial interest.

## ■ REFERENCES

- (1) 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.
- (2) Li, J.; Haycraft, C.; Iyengar, S. S. Hybrid, Extended Lagrangian – Born–Oppenheimer *Ab Initio* Molecular Dynamics using Fragment-Based Electronic Structure. *J. Chem. Theory Comput.* **2016**, *12*, 2493.
- (3) 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.
- (4) 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.
- (5) 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.
- (6) Schlegel, H. B.; Millam, J. M.; Iyengar, S. S.; Voth, G. A.; Daniels, A. D.; Scuseria, G. E.; Frisch, M. J. *Ab Initio* Molecular Dynamics: Propagating the Density Matrix with Gaussian Orbitals. *J. Chem. Phys.* **2001**, *114*, 9758.
- (7) Iyengar, S. S.; Schlegel, H. B.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J. *Ab Initio* Molecular Dynamics: Propagating the Density Matrix with Gaussian Orbitals. II. Generalizations Based on Mass-Weighting, Idempotency, Energy Conservation and Choice of Initial Conditions. *J. Chem. Phys.* **2001**, *115*, 10291.
- (8) Schlegel, H. B.; Iyengar, S. S.; Li, X.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J. *Ab Initio* Molecular Dynamics: Propagating the Density Matrix with Gaussian Orbitals. III. Comparison with Born–Oppenheimer Dynamics. *J. Chem. Phys.* **2002**, *117*, 8694.
- (9) Iyengar, S. S.; Schlegel, H. B.; Voth, G. A.; Millam, J. M.; Scuseria, G. E.; Frisch, M. J. *Ab Initio* Molecular Dynamics: Propagating the Density Matrix with Gaussian Orbitals. IV. Formal Analysis of the Deviations from Born–Oppenheimer Dynamics. *Isr. J. Chem.* **2002**, *42*, 191.
- (10) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. The ONIOM Method: Its Foundation and Applications to Metalloenzymes and Photobiology. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 327.
- (11) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643.

- (12) 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.
- (13) Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.* **2015**, *115*, 5607.
- (14) 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.
- (15) Yang, W. Direct Calculation of Electron Density in Density-Functional Theory. *Phys. Rev. Lett.* **1991**, *66*, 1438.
- (16) 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.
- (17) Kerdcharoen, T.; Morokuma, K. ONIOM-XS: An Extension of the ONIOM Method for Molecular Simulation in Condensed Phase. *Chem. Phys. Lett.* **2002**, *355*, 257.
- (18) 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.
- (19) Gordon, M.; Mullin, J.; Pruitt, S.; Roskop, L.; Slipchenko, L.; Boatz, J. Accurate Methods for Large Molecular Systems. *J. Phys. Chem. B* **2009**, *113*, 9646.
- (20) 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.
- (21) 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.
- (22) 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.
- (23) 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.
- (24) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many Body Expansion for Simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1.
- (25) 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.
- (26) 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.
- (27) 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.
- (28) 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.
- (29) Mayhall, N. J.; Raghavachari, K. Molecules-In-Molecules: An Extrapolated Fragment-Based Approach for Accurate Calculations on Large Molecules and Materials. *J. Chem. Theory Comput.* **2011**, *7*, 1336.
- (30) 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.
- (31) 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.
- (32) 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.
- (33) 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.
- (34) Kamiya, M.; Hirata, S.; Valiev, M. Fast Electron-Correlation Methods for Molecular Crystals Without Basis Set Superposition Errors. *J. Chem. Phys.* **2008**, *128*, 074103.
- (35) 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.
- (36) 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.
- (37) 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.
- (38) 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.
- (39) Björklund, A.; Husfeldt, T.; Koivisto, M. Set Partitioning via Inclusion Exclusion. *SIAM J. Comput.* **2009**, *39*, 546.
- (40) Sahu, N.; Yeole, S. D.; Gadre, S. R. Appraisal of molecular tailoring approach for large clusters. *J. Chem. Phys.* **2013**, *138*, 104101.
- (41) 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.
- (42) Hopkins, B. W.; Tschumper, G. S. Multicentred QM/QM Methods for Overlapping Model Systems. *Mol. Phys.* **2005**, *103*, 309.
- (43) Thapa, B.; Beckett, D.; Jovan Jose, K.; Raghavachari, K. Assessment of fragmentation strategies for large proteins using the multilayer molecules-in-molecules approach. *J. Chem. Theory Comput.* **2018**, *14*, 1383.
- (44) Kitaura, K.; Ikey, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701.
- (45) Collins, M. A. Systematic Fragmentation of Large Molecules by Annihilation. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7744.
- (46) Wen, S.; Nanda, K.; Huang, Y.; Beran, G. J. Practical quantum mechanics-based fragment methods for predicting molecular crystal properties. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7578.
- (47) Li, J.; Sode, O.; Hirata, S. Second-order many-body perturbation study on thermal expansion of solid carbon dioxide. *J. Chem. Theory Comput.* **2015**, *11*, 224.
- (48) 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.
- (49) Liu, J.; Zhu, T.; Wang, X.; He, X.; Zhang, J. Z. Quantum fragment based ab initio molecular dynamics for proteins. *J. Chem. Theory Comput.* **2015**, *11*, 5897.
- (50) Pruitt, S. R.; Nakata, H.; Nagata, T.; Mayes, M.; Alexeev, Y.; Fletcher, G.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. Importance of Three-Body interactions in molecular dynamics simulations of water demonstrated with the fragment molecular orbital method. *J. Chem. Theory Comput.* **2016**, *12*, 1423.
- (51) Collins, M. A. Can Systematic Molecular Fragmentation Be Applied to Direct Ab Initio Molecular Dynamics? *J. Phys. Chem. A* **2016**, *120*, 9281.
- (52) Liu, J.; He, X.; Zhang, J. Z.; Qi, L.-W. Hydrogen-bond structure dynamics in bulk water: insights from ab initio simulations with coupled cluster theory. *Chem. Sci.* **2018**, *9*, 2065–2073.
- (53) Varandas, A. J.; Murrell, J. N. A many-body expansion of polyatomic potential energy surfaces: application to H<sub>n</sub> systems. *Faraday Discuss. Chem. Soc.* **1977**, *62*, 92.
- (54) Pomes, R.; Roux, B. Structure and Dynamics of a Proton Wire: A Theoretical Study of H<sup>+</sup> Translocation Along the Single-File Water Chain in the Gramicidin a Channel. *Biophys. J.* **1996**, *71*, 19.

- (55) Brewer, M. L.; Schmitt, U. W.; Voth, G. A. The Formation and Dynamics of Proton Wires in Channel Environments. *Biophys. J.* **2001**, *80*, 1691.
- (56) Roux, B. Computational Studies of the Gramicidin Channel. *Acc. Chem. Res.* **2002**, *35*, 366.
- (57) Sagnella, D. E.; Laasonen, K.; Klein, M. L. *Ab Initio* molecular dynamics study of proton transfer in a polyglycine analog of the ion channel gramicidin A. *Biophys. J.* **1996**, *71*, 1172.
- (58) Tuckerman, M. E.; Laasonen, K.; Sprik, M.; Parrinello, M. *Ab Initio* Molecular Dynamics Simulation of the Solvation and Transport of  $\text{H}_3\text{O}^+$  and  $\text{OH}^-$ . *J. Phys. Chem.* **1995**, *99*, 5749.
- (59) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. The Nature of the Hydrated Excess Proton in Water. *Nature* **1999**, *397*, 601.
- (60) Isaacs, N. *Physical Organic Chemistry*; Longman Scientific and Technical, Essex, 1995.
- (61) Sheridan, R. Quantum Mechanical Tunneling in Organic Reactive Intermediates. In *Reviews of Reactive Intermediate Chemistry*; Platz, M., Moss, R. A., Jones, M., Jr., Eds.; Wiley-Interscience: Hoboken, NJ, USA, 2007.
- (62) Dey, T. K.; Shah, N. R. On the number of simplicial complexes in  $\mathbb{R}^d$ . *Comput. Geom.* **1997**, *8*, 267.
- (63) Adams, C. C.; Franzosa, R. D. *Introduction to topology: Pure and applied*; 2008.
- (64) Mecke, K. R.; Wagner, H. Euler characteristic and related measures for random geometric sets. *J. Stat. Phys.* **1991**, *64*, 843.
- (65) Bowyer, A. Computing Dirichlet tessellations. *Comput. J.* **1981**, *24*, 162.
- (66) Watson, D. F. Computing the n-dimensional Delaunay tessellation with applications to Voronoi polytopes. *Comput. J.* **1981**, *24*, 167.
- (67) Aurenhammer, F. Voronoi Diagrams — A survey of a fundamental geometric data structure. *ACM Comput. Survey* **1991**, *23*, 345.
- (68) Okabe, A.; Boots, B.; Sugihara, K.; Chiu, S. N. *Spatial Tessellations — Concepts and applications of Voronoi diagrams*; John Wiley and Sons: Chichester, U.K., 2000; DOI: 10.1002/9780470317013.
- (69) Hert, S.; Seel, M. dD Convex Hulls and Delaunay Triangulations. *CGAL User and Reference Manual*, 4th ed.; CGAL Editorial Board, 2018.
- (70) Farin, G. Surfaces over Dirichlet Tessellations. *Comput. Aided Geom. Des.* **1990**, *7*, 281.
- (71) Armstrong, M. A. *Basic topology*; Springer Science & Business Media: New York, 2013.
- (72) Edelsbrunner, H. *Geometry and topology for mesh generation*; Cambridge University Press: New York, 2001; Vol. 7, DOI: 10.1017/CBO9780511530067.
- (73) Flegg, H. G. *From Geometry to Topology*; Dover: Mineola, NY, USA, 2001.
- (74) Levine, N. Semi-open sets and semi-continuity in topological spaces. *Am. Math. Mon.* **1963**, *70*, 36.
- (75) Padlewska, B.; Darmochwał, A. Topological spaces and continuous functions. *Formalized Math.* **1990**, *1*, 223.
- (76) 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.
- (77) 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.
- (78) 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 Water. *J. Chem. Phys.* **2002**, *117*, 5839.
- (79) Baer, M. *Beyond Born-Oppenheimer: Conical Intersections and Electronic Nonadiabatic Coupling Terms*; John Wiley and Sons: New York, 2006.
- (80) Hanna, G.; Kapral, R. Nonadiabatic Dynamics of Condensed Phase Rate Processes. *Acc. Chem. Res.* **2006**, *39*, 21.
- (81) Worth, G. A.; Cederbaum, L. S. Beyond Born-Oppenheimer: Molecular Dynamics Through a Conical Intersection. *Annu. Rev. Phys. Chem.* **2004**, *55*, 127.
- (82) Jasper, A. W.; Zhu, C.; Nangia, S.; Truhlar, D. G. Introductory Lecture: Nonadiabatic Effects in Chemical Dynamics. *Faraday Discuss.* **2004**, *127*, 1.
- (83) Kendrick, B. K.; Alden Mead, C.; Truhlar, D. G. Properties of Nonadiabatic Couplings and the Generalized Born Oppenheimer Approximation. *Chem. Phys.* **2002**, *277*, 31.
- (84) 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, New York, 1996; p 411.
- (85) Yarkony, D. R. Diabolical Conical Intersections. *Rev. Mod. Phys.* **1996**, *68*, 985.
- (86) 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.
- (87) Lin, H.; Zhao, Y.; Tishchenko, O.; Truhlar, D. G. Multi-configuration Molecular Mechanics Based on Combined Quantum Mechanical and Molecular Mechanical Calculations. *J. Chem. Theory Comput.* **2006**, *2*, 1237.
- (88) Mulliken, R. S. The Interaction of Electron Donors and Acceptors. *J. Chim. Phys. Phys.-Chim. Biol.* **1964**, *61*, 20.
- (89) Borgis, D.; Staib, A. A Semiempirical Quantum Polarization Model for Water. *Chem. Phys. Lett.* **1995**, *238*, 187.
- (90) Heyden, A.; Lin, H.; Truhlar, D. G. Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations. *J. Phys. Chem. B* **2007**, *111*, 2231.
- (91) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. An improvement in quantum mechanical description of solute-solvent interactions in condensed systems via the number-adaptive multiscale quantum mechanical/molecular mechanical-molecular dynamics method: Application to zwitterionic glycine in aqueous solution. *J. Chem. Phys.* **2012**, *137*, 024501.
- (92) Fleurat-Lessard, P.; Michel, C.; Buló, R. E. Energy extrapolation schemes for adaptive multi-scale molecular dynamics simulations. *J. Chem. Phys.* **2012**, *137*, 074111.
- (93) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. An improvement in quantum mechanical description of solute-solvent interactions in condensed systems via the number-adaptive multiscale quantum mechanical/molecular mechanical-molecular dynamics method: Application to zwitterionic glycine in aqueous solution. *J. Chem. Phys.* **2012**, *137*, 024501.
- (94) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. Effective force fields for condensed phase systems from *ab initio* molecular dynamics simulation: A new method for force-matching. *J. Chem. Phys.* **2004**, *120*, 10896.
- (95) Coker, D. F. In *Computer Simulation in Chemical Physics*; Allen, M. P., Tildesley, D. J., Eds.; Kluwer Academic: Dordrecht, The Netherlands, 1993.
- (96) Mead, C. A.; Truhlar, D. G. On the Determination of Born-Oppenheimer Nuclear Motion Wave Functions Including Complications Due to Conical Intersections and Identical Nuclei. *J. Chem. Phys.* **1979**, *70*, 2284.
- (97) Berry, M. V. Quantal Phase Factors Accompanying Adiabatic Changes. *Proc. R. Soc. London, Ser. A* **1984**, *392*, 45.
- (98) Tully, J. C.; Preston, R. K. Trajectory surface hopping approach to nonadiabatic molecular collisions: the reaction of  $\text{H}^+$  with  $\text{D}_2$ . *J. Chem. Phys.* **1971**, *55*, 562.
- (99) Blondel, V. D.; Gajardo, A.; Heymans, M.; Senellart, P.; Van Dooren, P. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM Rev.* **2004**, *46*, 647.
- (100) MacQueen, J. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. 1967; Vol. 1, p 281.
- (101) Daubechies, I. *Ten Lectures in Wavelets*; SIAM: Philadelphia, PA, USA, 1992; DOI: 10.1137/1.9781611970104.

- (102) Strang, G.; Nguyen, T. *Wavelets and Filter Banks*, 2nd ed.; Wellesley-Cambridge Press, Wellesley, MA, USA, 1996.
- (103) Strang, G.; Strela, V. Orthogonal Multiwavelets with Vanishing Moments. *Opt. Eng.* **1994**, *33*, 2104.
- (104) Arias, T. Multiresolution Analysis of Electronic Structure: Semicardinal and Wavelet Bases. *Rev. Mod. Phys.* **1999**, *71*, 267.
- (105) Johnson, B. R.; Modiset, J. P.; Nordlander, P. J.; Kinsey, J. L. Quadrature Integration for Orthogonal Wavelet Systems. *J. Chem. Phys.* **1999**, *110*, 8309.
- (106) Chan, Y. T. *Wavelet Basics*; Kluwer Academic, Norwell, MA, USA, 1995.
- (107) Strang, G. On the Construction and Comparison of Difference Schemes. *SIAM J. Numer. Anal.* **1968**, *5*, 506.
- (108) 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.
- (109) 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.
- (110) 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.
- (111) Hocker, D.; Li, X.; Iyengar, S. S. Shannon Information Entropy Based Time-Dependent Deterministic Sampling Techniques for Efficient “on-The-Fly” Quantum Dynamics and Electronic Structure. *J. Chem. Theory Comput.* **2011**, *7*, 256.
- (112) Li, J.; Li, X.; Iyengar, S. S. Vibrational Properties of Hydrogen Bonded Systems Using the Multi-Reference Generalization to the “On-The-Fly” Electronic Structure Within Quantum Wavepacket Ab Initio Molecular Dynamics (QWAIMD). *J. Chem. Theory Comput.* **2014**, *10*, 2265.
- (113) 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.
- (114) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian16*, Revision B.01; Gaussian: Wallingford, CT, USA, 2016.
- (115) Pomes, R.; Roux, B. Theoretical Study of H<sup>+</sup> Translocation Along a Model Proton Wire. *J. Phys. Chem.* **1996**, *100*, 2519.
- (116) Schmitt, U. W.; Voth, G. A. The Computer Simulation of Proton Transport in Water. *J. Chem. Phys.* **1999**, *111*, 9361.
- (117) Sadeghi, R. R.; Cheng, H.-P. The Dynamics of Proton Transfer in a Water Chain. *J. Chem. Phys.* **1999**, *111*, 2086.
- (118) Nagle, J. F.; Morowitz, H. J. Molecular Mechanisms for Proton Transport in Membranes. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 298.
- (119) 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.
- (120) Guo, H.; Barnard, A. S. Proton transfer in the hydrogen-bonded chains of lepidocrocite: a computational study. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17864.
- (121) Domene, C.; Sansom, M. S. Potassium channel, ions, and water: simulation studies based on the high resolution X-ray structure of KcsA. *Biophys. J.* **2003**, *85*, 2787.
- (122) Dellago, C.; Naor, M. M.; Hummer, G. Proton transport through water-filled carbon nanotubes. *Phys. Rev. Lett.* **2003**, *90*, 105902.
- (123) Mann, D. J.; Halls, M. D. Water alignment and proton conduction inside carbon nanotubes. *Phys. Rev. Lett.* **2003**, *90*, 195503.
- (124) Cao, Z.; Peng, Y.; Yan, T.; Li, S.; Li, A.; Voth, G. A. Mechanism of fast proton transport along one-dimensional water chains confined in carbon nanotubes. *J. Am. Chem. Soc.* **2010**, *132*, 11395.
- (125) Ye, Y.-S.; Rick, J.; Hwang, B.-J. Water soluble polymers as proton exchange membranes for fuel cells. *Polymers* **2012**, *4*, 913.
- (126) Ermiler, U.; Fritzsche, G.; Buchanan, S. K.; Michel, H. Structure of the photosynthetic reaction centre from Rhodobacter sphaeroides at 2.65 Å resolution: cofactors and protein-cofactor interactions. *Structure* **1994**, *2*, 925.
- (127) Iyengar, S. S.; Day, T. J. F.; Voth, G. A. On the Amphiphilic Behavior of the Hydrated Proton: An Ab Initio Molecular Dynamics Study. *Int. J. Mass Spectrom.* **2005**, *241*, 197.
- (128) Iyengar, S. S. Further Analysis of the Dynamically Averaged Vibrational Spectrum for the “magic” Protonated 21-Water Cluster. *J. Chem. Phys.* **2007**, *126*, 216101.
- (129) Li, X.; Moore, D. T.; Iyengar, S. S. Insights from First Principles Molecular Dynamics Studies Towards Infra-Red Multiple-Photon and Single-Photon Action Spectroscopy: Case Study of the Proton-Bound Di-Methyl Ether Dimer. *J. Chem. Phys.* **2008**, *128*, 184308.
- (130) Zundel, G. In *The Hydrogen Bond: Recent Developments in Theory and Experiments. II. Structure and Spectroscopy*; Schuster, P., Zundel, G., Sandorfy, C., Eds.; North-Holland: Amsterdam, 1976; p 683.
- (131) 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.
- (132) 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<sub>3</sub>O<sub>2</sub><sup>-</sup> and H<sub>5</sub>O<sub>2</sub><sup>+</sup> Complexes. *J. Phys. Chem. A* **2005**, *109*, 1487.
- (133) Vendrell, O.; Gatti, F.; Meyer, H.-D. Dynamics and Infrared Spectroscopy of the Protonated Water Dimer. *Angew. Chem., Int. Ed.* **2007**, *46*, 6918.
- (134) Dietrick, S. M.; Iyengar, S. S. Constructing Periodic Phase Space Orbits from Ab Initio Molecular Dynamics Trajectories to Analyze Vibrational Spectra: Case Study of the Zundel (H<sub>5</sub>O<sub>2</sub><sup>+</sup>) Cation. *J. Chem. Theory Comput.* **2012**, *8*, 4876.
- (135) Warshel, A.; Papazyan, A.; Kollman, P. A. On Low-Barrier Hydrogen-Bonds and Enzyme Catalysis. *Science* **1995**, *269*, 102.
- (136) Perrin, C. L.; Nielson, J. B. ‘Strong’ Hydrogen Bonds in Chemistry and Biology. *Annu. Rev. Phys. Chem.* **1997**, *48*, 511.
- (137) Eigen, M.; Maeyer, L. D. Self-Dissociation and Protonic Charge Transport in Water and Ice. *Proc. R. Soc. London, Ser. A* **1958**, *247*, 505.
- (138) Eigen, M. Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. Part I: Elementary Processes. *Angew. Chem., Int. Ed. Engl.* **1964**, *3*, 1.
- (139) Bergmann, U.; Di Cicco, A.; Wernet, P.; Principi, E.; Glatzel, P.; Nilsson, A. Nearest-neighbor oxygen distances in liquid water and ice observed by x-ray Raman based extended x-ray absorption fine structure. *J. Chem. Phys.* **2007**, *127*, 174504.
- (140) Hynes, J. T., Klinman, J. P., Limbach, H.-H., Schowen, R. L., Eds. *Hydrogen-Transfer Reactions*; Wiley-VCH: Weinheim, Germany, 2007; DOI: 10.1002/9783527611546.

- (141) Nagel, Z.; Klinman, J. Tunneling and Dynamics in Enzymatic Hydride Transfer. *Chem. Rev.* **2006**, *106*, 3095.
- (142) Rickert, K. W.; Klinman, J. P. Nature of Hydrogen Transfer in Soybean Lipoxygenase 1: Separation of Primary and Secondary Isotope Effects. *Biochemistry* **1999**, *38*, 12218.
- (143) Peeters, J.; Nguyen, T. L.; Vereecken, L. HOx Radical Regeneration in the Oxidation of Isoprene. *Phys. Chem. Chem. Phys.* **2009**, *11*, 5935.
- (144) Abou-Zied, O. K. Examining [2, 2-bipyridyl]-3, 3-diol as a possible DNA model base pair. *J. Photochem. Photobiol., A* **2006**, *182*, 192.
- (145) Phatak, P.; Sumner, I.; Iyengar, S. S. Gauging the Flexibility of the Active Site in Soybean Lipoxygenase-1 (SLO-1) Through an Atom-Centered Density Matrix Propagation (ADMP) Treatment That Facilitates the Sampling of Rare Events. *J. Phys. Chem. B* **2012**, *116*, 10145.
- (146) Phatak, P.; Venderley, J.; Debrot, J.; Li, J.; Iyengar, S. S. Active Site Dynamical Effects That Facilitate the Hydrogen Transfer Process in Soybean Lipoxygenase-1 (SLO-1): Isotope Effects. *J. Phys. Chem. B* **2015**, *119*, 9532.
- (147) Sepiol, J.; Bulska, H.; Grabowska, A. The dihydroxy derivative of 2, 2'-bipyridyl as a new proton-transfer lasing dye. *Chem. Phys. Lett.* **1987**, *140*, 607.
- (148) Kaczmarek, Ł.; Borowicz, P.; Grabowska, A. Strongly modified [2, 2-bipyridyl]-3, 3-diol (BP (OH) 2): a system undergoing excited state intramolecular proton transfer as a photostabilizer of polymers and as a solar energy collector. *J. Photochem. Photobiol., A* **2001**, *138*, 159.
- (149) Haile, S. M.; Boysen, D. A.; Chisholm, C. R. I.; Merle, R. B. Solid Acids As Fuel Cell Electrolytes. *Nature* **2001**, *410*, 910.
- (150) Knox, C. K.; Voth, G. A. Probing Selected Morphological Models of Hydrated Nation Using Large-Scale Molecular Dynamics Simulations. *J. Phys. Chem. B* **2010**, *114*, 3205.
- (151) Shida, N.; Barbara, P. F.; Almlöf, J. A reaction surface Hamiltonian treatment of the double proton transfer of formic acid dimer. *J. Chem. Phys.* **1991**, *94*, 3633.
- (152) Chocholoušová, J.; Vacek, J.; Hobza, P. Acetic acid dimer in the gas phase, nonpolar solvent, microhydrated environment, and dilute and concentrated acetic acid: *ab initio* quantum chemical and molecular dynamics simulations. *J. Phys. Chem. A* **2003**, *107*, 3086.
- (153) Kim, Y. Direct dynamics calculation for the double proton transfer in formic acid dimer. *J. Am. Chem. Soc.* **1996**, *118*, 1522.
- (154) Ishikawa, H.; Iwata, K.; Hamaguchi, H.-o. Picosecond dynamics of stepwise double proton-transfer reaction in the excited state of the 2-aminopyridine/acetic acid system. *J. Phys. Chem. A* **2002**, *106*, 2305.
- (155) Hayashi, S.; Umemura, J.; Kato, S.; Morokuma, K. *Ab initio* molecular orbital study on the formic acid dimer. *J. Phys. Chem.* **1984**, *88*, 1330.
- (156) Chang, Y. T.; Yamaguchi, Y.; Miller, W. H.; Schaefer, H. F., III An analysis of the infrared and Raman spectra of the formic acid dimer (HCOOH) 2. *J. Am. Chem. Soc.* **1987**, *109*, 7245.
- (157) Miura, S.; Tuckerman, M. E.; Klein, M. L. An *ab initio* path integral molecular dynamics study of double proton transfer in the formic acid dimer. *J. Chem. Phys.* **1998**, *109*, 5290.
- (158) Qu, C.; Bowman, J. M. An *ab initio* potential energy surface for the formic acid dimer: zero-point energy, selected anharmonic fundamental energies, and ground-state tunneling splitting calculated in relaxed 1–4-mode subspaces. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24835.
- (159) Luckhaus, D. Hydrogen exchange in formic acid dimer: tunnelling above the barrier. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8357.
- (160) Gao, J.; Amara, P.; Alhambra, C.; Field, M. A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A* **1998**, *102*, 4714.
- (161) Humbel, S.; Sieber, S.; Morokuma, K. The IMOMO Method: Integration of Different Levels of Molecular Orbital Approximations for Geometry Optimization of Large Systems: Test for N-Butane Conformation and SN<sub>2</sub> Reaction: RCl+Cl<sup>-</sup>. *J. Chem. Phys.* **1996**, *105*, 1959.
- (162) Aurenhammer, F. Voronoi diagrams a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)* **1991**, *23*, 345.
- (163) Okabe, A.; Boots, B.; Sugihara, K.; Chiu, S. N. *Spatial tessellations: concepts and applications of Voronoi diagrams*; John Wiley & Sons: Chichester, U.K., 2000; Vol. 501.
- (164) Coffey, T. M.; Wyatt, R. E.; Schieve, W. C. Reconstruction of the time-dependent wave function exclusively from position data. *Phys. Rev. Lett.* **2011**, *107*, 230403.