

Large Language Model-Type Architecture for High-Dimensional Molecular Potential Energy Surfaces

Xiao Zhu¹ and Srinivasan S. Iyengar^{1*}

*Department of Chemistry, Department of Physics, and the Indiana University Quantum Science and Engineering Center (IU-QSEC), Indiana University,
800 East Kirkwood Avenue, Bloomington, Indiana 47405, USA*



(Received 10 December 2024; revised 10 September 2025; accepted 5 January 2026; published 22 January 2026)

Computing high-dimensional potential energy surfaces for molecular systems and materials is considered to be a great challenge in computational chemistry with potential impact in a range of areas including the fundamental prediction of reaction rates. In this paper, we design and discuss an algorithm that has similarities to large language models in generative AI and natural language processing. Specifically, we represent a molecular system as a graph which contains a set of nodes, edges, faces, etc. Interactions between these sets, which represent molecular subsystems in our case, are used to construct the potential energy surface for a reasonably sized chemical system with 51 nuclear dimensions. For this purpose, a family of neural networks that pertain to the graph-theoretically obtained subsystems get the job done for this 51 nuclear dimensional system. We then ask if this same family of lower-dimensional graph-based neural networks can be transformed to provide accurate predictions for a 186-dimensional potential energy surface. We find that our algorithm does provide accurate results for this larger-dimensional problem with sub-kcal/mol accuracy for the higher-dimensional potential energy surface problem. Indeed, as a result of these developments, here we produce efforts toward a full-dimensional potential energy surface for the protonated 21-water cluster (186 nuclear dimensions) at coupled-cluster single and double level accuracy.

DOI: [10.1103/2qcy-8n8g](https://doi.org/10.1103/2qcy-8n8g)

Subject Areas: Atomic and Molecular Physics,
Chemical Physics, Physical Chemistry

I. INTRODUCTION

The problem of computing accurate post-Hartree-Fock potential energy surfaces for large chemical systems is considered an exponentially complex task in quantum chemistry. However, computing such accurate potentials is essential for a range of problems that include not only the fundamental description of chemical reactions and molecular vibrations beyond the harmonic approximation [1–8], but also topics such as the design of new molecular assemblies with desired properties [9–13]. The complexity of describing these phenomena is governed by the following fundamental limitations. As the system size grows, the number of nuclear dimensions increases. For $\mathcal{D}$ nuclear dimensions with each dimension discretized and represented using M multidimensional basis points, where such basis points may represent nuclear geometries, the number

of such geometries grows exponentially with the number of nuclear dimensions, and the problem may scale as $O(M^{\mathcal{D}})$ [1–3,6,14–23]. Additionally, if post-Hartree-Fock accuracy (such as coupled-cluster theory) is desired, computing each potential energy may scale as $O(N^{7-10})$, for N electrons [24]. Thus, the problem of computing accurate potential energy surfaces grows as $O(M^{\mathcal{D}} * N^{7-10})$, which becomes intractable even for moderate-sized systems.

Recently, many groups have used machine-learning methods to compute accurate potential energy surfaces [25–34]. However, for obtaining multidimensional potentials, the amount of data needed to train these models could also grow steeply, and, additionally, when accurate electron correlation information is needed, as noted above, the training cost may be prohibitive. The complexity of obtaining accurate neural network (NN) models may be understood from the density of the edges on the right side in Fig. 1. Here, the multidimensional electronic potential energy surface for a simple water-wire system, depicted as a graphical chain with individual water molecules shown as nodes, is extrapolated using the neural network on the right. The gray edges represent the neural network weights that need to be computed through training, and the density of such edges represents the complexity of the training problem. Figure 1

*Contact author: iyengar@iu.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

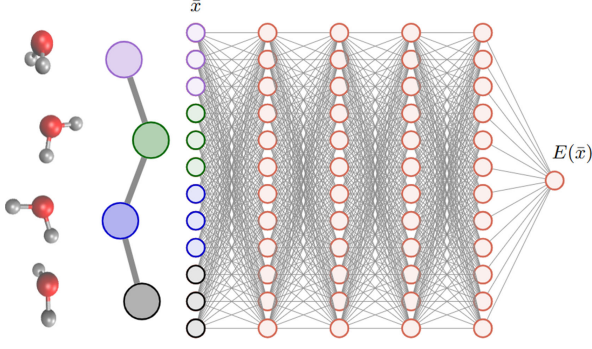


FIG. 1. The density of gray edges represents the number of network weights that need to be computed to obtain a representation of the potential energy surface.

does not account for the steep algebraic scaling needed to compute the training data (which is generally represented by the span of the nuclear coordinates variable, $\bar{x}$ in Fig. 1) within the training set. Because of such reasons, the largest system studied to date using coupled-cluster singles and doubles with perturbative triples [CCSD(T)] data is H_9O_4^+ [27] (with an absolute error of 0.09 kcal/mol). Furthermore, the resultant neural network model shows limited transferability, and the mean absolute error for $\text{H}_{13}\text{O}_6^+$ using this model is found to be 1.32 kcal/mol [25,35,36]. It is precisely this transferability problem that we address in this paper. In contrast to the above-listed studies, the machine-learning methods developed here result in 0.36 kcal/mol accuracy for the coupled-cluster single and double (CCSD) potential energy surface for the complex protonated 21-water cluster ($\text{H}_{64}\text{O}_{21}^+$) system.

Toward this effort, in this paper, we introduce a machine-learning-based approach to compute multidimensional potential energy surfaces that we show is applicable for much larger systems than previously described. The goals that we address in this paper are as follows: (i) We first ask if it is possible to construct a general formalism for accurate multidimensional potential energy surfaces for moderately sized systems using neural networks that are far less dense and require far less training than implied by Fig. 1. We propose an approach for this purpose using graph-theoretic protocols. (ii) Once problem (i) is addressed reasonably well, we then wonder if it is possible to use the knowledge from these neural networks to then develop a solution for a potential energy surface size that is beyond what can currently be done. (iii) In solving the problem in (ii), we devise a method that we find to be related, in a very fundamental way, to the concept of “attention” in large language model methods [37] and generative AI.

More specifically, the first key question we attempt to address here, as part of item (i) above, is whether neural networks of lower complexity, that is, neural networks with fewer degrees of freedom corresponding to systems requiring fewer learning cycles and training effort, can be rigorously harnessed together to create a larger network

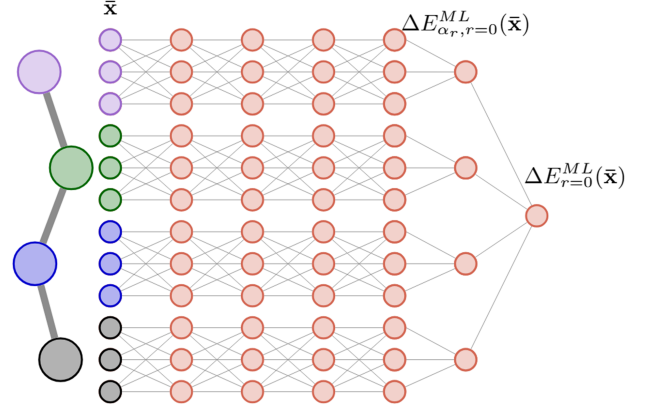


FIG. 2. Visual illustration of neural networks used to compute $\{\Delta E_{\alpha,r}^{ML}(\bar{x})\}$ for $r = 0$. See Eqs. (2), (7), and (9).

that can be used to solve a larger problem. Toward this, we use our graph-theoretic procedure for molecular fragmentation [38–54]. This approach may be thought of as a generalized convolution neural network [55,56], where the convolution is not uniform and is dictated by the spatial proximity of groups within a molecular system and is defined through a graphical network. An example of how such a scheme works is shown in Figs. 2 and 3, where the neural network in Fig. 1 is visually simplified into a family of networks defined by a graphical representation of the molecular system. Note that Fig. 2 results in four separate neural networks, and Fig. 3 results in three separate neural networks, each of which is much lower in complexity as compared to Fig. 1. These separate neural networks are then combined using the topology of the graph description. In Refs. [46,49], we have shown that this approach can be

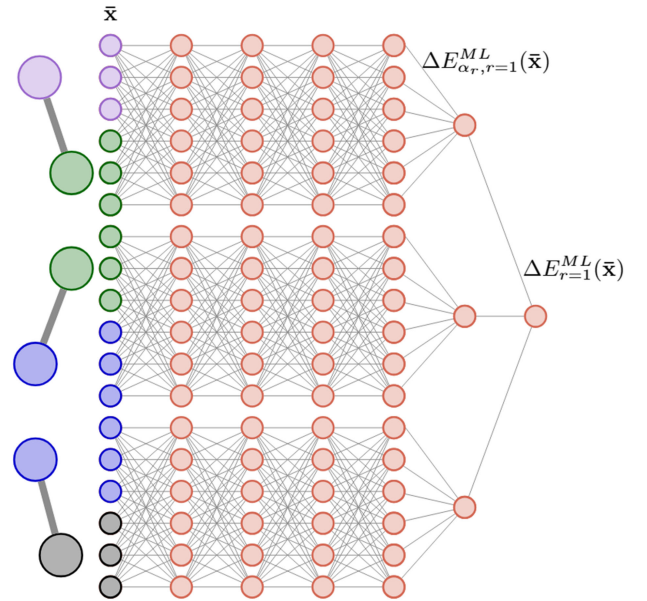


FIG. 3. Visual illustration of neural networks used to compute $\{\Delta E_{\alpha,r}^{ML}(\bar{x})\}$ for $r = 1$. See Eqs. (2), (7), and (9).

used as a graph-theory-based transfer learning protocol where fragment neural networks can lead to accurate potential energy surfaces in higher dimensions [46,49].

Then, as part of item (ii) above, we show that the individual models of these graph-based molecular fragments can be incrementally further improved to provide accurate results for systems even larger than those for which they were originally parametrized. Specifically, in Ref. [49], we showed how neural networks written for small water clusters such as protonated or neutral dimers and trimers can be used to compute accurate potentials with errors in the sub-kcal/mol range for a larger solvated Zundel system. Here, as part of item (ii) above, we show how these same sets of neural networks can be further refined to obtain sub-kcal/mol accuracy for potential energy surfaces for even larger systems such as the protonated 21-water cluster system. Thus, in this paper, for the first time, we provide a systematic strategy to obtain accurate, coupled-cluster-level potentials for large-dimensional systems. Here, we illustrate our procedure for a 186-dimensional problem, the protonated 21-water cluster system. Finally, as part of item (iii) above, we discover that our approach here is deeply related to the concept of attention in generative AI and large-language models [37].

This paper is organized as follows: In Sec. II, we introduce our graph-theory approach for accurate post-Hartree-Fock electronic structure potential energies; then, in Sec. II A, we show how this approach yields a method for computing electronic energies by introducing a family of neural networks with far-reduced computational complexity. Training of the resultant neural networks is carried out using a geometric tessellation of high-dimensional potential energy surface spaces, and this method is described in Appendix B. This geometric tessellation approach allows us to train neural networks with only 10% training data, and the remaining 90% data are used for testing and validation. This is a critical aspect of our approach, because, if coupled-cluster accuracy is needed [scaling: $\mathcal{O}(N^{6\cdots 7})$], the training data must be as compressed as possible. In Sec. III, we probe the connections between our graph-theoretic methodology and the concept of attention in “transformers” used in large language models (LLMs) and generative AI. As discussed later, our method has key differences from the well-known graph attention network architecture [57]. While graph attention networks apply learnable attention mechanisms to graph components via neural-network-based embeddings, the graph weights in our formalism that lend naturally to an interpretation as attention in our approach are defined deterministically, based on the topology and connectivity of the graph. Furthermore, the graph weights used here include all simplex ranks, that is, include nodes, edges, faces, and higher-order molecular interactions tempered by their respective geometric weights. As we discuss in Sec. II, the weights for each graph component are computed

directly from geometric and topological considerations, no learnable parameters are introduced for the embedding construction, and no optimization is needed to obtain or refine these attention scores.

Our numerical tests are conducted on water clusters. Water clusters have received wide attention in the literature due to their broad and fundamental significance [58–68] and also due to challenges presented to both experiment and theory [59–64,66]. The challenges arise from the need for multidimensional quantum-mechanical treatment of the light hydrogen nuclear degrees of freedom (that is, span and complexity of the potential energy surface), along with the need for potential energy surfaces with post-Hartree-Fock (electron-correlation) accuracy. Thus, using the approaches developed here, we show that we are able to generate accurate potential energy surfaces for the solvated Zundel ($\text{H}_{13}\text{O}_6^+$) system (51 nuclear dimensions), and these results are discussed in Sec. IV. Additional technical aspects are discussed in Appendixes B and C. The next question addressed in this paper is whether the family of neural networks which was proved to be sufficient for the solvated Zundel system in Sec. IV remains sufficient for the much larger and more complex protonated 21-water cluster system. The protonated 21-water cluster system, as we discuss, has presented a great challenge for experiment and theory due to the large number of quantum nuclear degrees of freedom and associated complexity of potential energy surface evaluations. We find, in Sec. IV, that the potential energy surface for the protonated 21-water cluster is such that the neural networks computed from the solvated Zundel system are not sufficient to capture the surface accurately. We, thus, present an incremental tessellation scheme to improve the neural network models that were used for solvated Zundel system which helps achieve sub-kcal/mol accuracy. Conclusions are given in Sec. VII.

II. GRAPH-THEORETIC METHODS TO REDUCE NEURAL NETWORK COMPLEXITY

In order to reduce the complexity of the network in Fig. 1 to that in Figs. 2 and 3, we first discuss a graph-based molecular fragmentation of our chemical system. This graph then becomes the basis for creating our convolution diagrams that leads to a family of decoupled neural networks. A graph $\mathcal{G}$, composed of molecular fragments treated as nodes (or vertices), is defined as $\mathcal{G} \equiv \{\mathbf{V}_0; \mathbf{V}_1\}$. Here, $\mathbf{V}_0$ is the set of vertices (fragments within a molecule or a cluster), and $\mathbf{V}_1$ is the set of edges that capture interactions between these vertices and, hence, are larger fragments. Once such a graph is defined, higher-rank objects, and, hence, larger fragments and molecular aggregates that represent high-order interactions up to a maximum rank $\mathcal{R}$, are also specified by the graph as

$$\{\mathbf{V}_r | r = 0, \dots, \mathcal{R}\} \equiv \{\mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_r, \dots, \mathbf{V}_{\mathcal{R}}\}. \quad (1)$$

The quantity $\mathbf{V}_r$ represents the set of all rank- r entities in the graph that may be used to capture the interactions between $(r+1)$ -rank molecular fragments.

The post-Hartree-Fock, “*target*,” electronic potential energy surface may then be written using such a graph-theoretic representation as

$$E^{\text{target}}(\bar{\mathbf{x}}) = E^{\text{Ref}}(\bar{\mathbf{x}}) + \sum_{r=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r, r}^{\mathcal{R}} \Delta E_{\alpha_r, r}(\bar{\mathbf{x}}), \quad (2)$$

where

$$\Delta E_{\alpha_r, r}(\bar{\mathbf{x}}) = E_{\alpha_r, r}^{\text{target}}(\bar{\mathbf{x}}) - E_{\alpha_r, r}^{\text{Ref}}(\bar{\mathbf{x}}). \quad (3)$$

Note here that the variable $\bar{\mathbf{x}}$ on the left side of Eq. (2) represents the full multidimensional molecular coordinate system corresponding to all nuclear degrees of freedom in the problem. On the right side of Eq. (2), however, only that portion of the system that is contained within the fragment labeled by the indices (α_r, r) needs to be included within the functional dependence of $\Delta E_{\alpha_r, r}(\bar{\mathbf{x}})$. That is, $\Delta E_{\alpha_r, r}(\bar{\mathbf{x}}) \equiv \Delta E_{\alpha_r, r}(\bar{\mathbf{x}}_{\alpha_r, r})$. This drastic reduction in dimensionality is similar to that seen in methods such as high-dimensional model representations [4,19,20,69–71]. Thus,

$$\begin{aligned} \Delta E^{\text{target} \leftarrow \text{Ref}}(\bar{\mathbf{x}}) &= E^{\text{target}}(\bar{\mathbf{x}}) - E^{\text{Ref}}(\bar{\mathbf{x}}) \\ &= \sum_{r=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r, r}^{\mathcal{R}} \Delta E_{\alpha_r, r}(\bar{\mathbf{x}}_{\alpha_r, r}). \end{aligned} \quad (4)$$

The designation “Ref” in the equations above is for some reference electronic structure level such as density-functional theory (DFT) at nuclear configuration $\bar{\mathbf{x}}$. In Eq. (2), a high-level energy correction $\Delta E_{\alpha_r, r}(\bar{\mathbf{x}})$ for each fragment or simplex, $\alpha_r \in \mathbf{V}_r$, is added to a reference energy, such as DFT. The quantity

$$\mathcal{M}_{\alpha_r, r} = \sum_{m \geq r} (-1)^{m+r} p_{\alpha_r}^{r, m} \quad (5)$$

represents the total number of times the simplex $\alpha_r \in \mathbf{V}_r$ appears in all simplexes with rank greater than r . Thus, $\mathcal{M}_{\alpha_r, r}$ is an overcounting correction, and $p_{\alpha_r}^{r, m}$ is the number of times the α_r^{th} rank- r simplex is included within all rank- m simplexes, for $m \geq r$.

Equation (2) has been widely benchmarked over a series of publications. Specifically, we have shown that Eq. (2) may be used to construct on-the-fly *ab initio* molecular dynamics (AIMD) trajectories [38–42] and potential energy surface calculations [43,47,48,51] for gas-phase as well as condensed-phase systems [44,46]. Both extended Lagrangian as well as Born-Oppenheimer-based *ab initio* molecular dynamics simulations can be performed at accuracy comparable to CCSD and second-order Møller-Plesset perturbation theory (MP2) levels of theory with

DFT-computational cost [38–41]. Hence, for the first time, in Refs. [39,40] we presented Car-Parrinello-style dynamics but with CCSD accuracy. Similarly, we have shown how multiple graphical representations of molecular systems can be used simultaneously to construct accurate potential energy surfaces in agreement with the MP2 and CCSD levels of theory, again at DFT cost [43,45,47,48,51]. In Ref. [42], we have also shown that weak interactions (specifically, hydrogen bonds) can be accurately captured and efficient approximations to large-basis AIMD trajectories, such as 6-311++G(2df,2pd), can be constructed through computational effort commensurate with much smaller basis set sizes, sets such as 6-31+G(d). Furthermore, we have also shown in Refs. [44,46] how condensed-phase simulations on interfaces and liquids may be constructed with hybrid DFT accuracy at gradient-corrected DFT accuracy. We also provide novel approaches to construct reduced circuit depth quantum computing algorithms in Refs. [50,53]. Finally, the approach has been demonstrated for hydrogen-bonded systems (such as water clusters and condensed phase systems) and also for covalently bonded biological systems and hydrocarbon where link atoms [41,42,46] are included within Eq. (2). This work is reviewed in Refs. [45,52,54].

Within the context of smooth molecular potential energy surfaces [43,45,47], we have shown how multiple graphical topologies can be combined into a single weighted graph formalism. Since the graphs are defined based on the instantaneous molecular geometry, the network encoded within the graphs (that help to capture local and nonlocal correlation and basis-set effects) may change when atoms move. This can happen when chemical bonds break or new ones are created. In Refs. [43,45,47], we show that this also happens when protons move in a water wire, such as that appearing in the gramicidin ion channel [72–74]. This may create singular hops in potential energy surfaces as atoms move. In Refs. [43,47], we have introduced a weighted graph approach to overcome these singularities. Given multiple graphs $\{\mathcal{G}_\beta\}$ for a molecular configuration $\bar{\mathbf{x}}$, with each graph yielding a single energy correction function: $(\bar{\mathbf{x}}, \mathcal{G}_\beta) \mapsto \Delta E_{\beta, \mathcal{R}_\beta}^{\text{target} \leftarrow \text{Ref}}(\bar{\mathbf{x}}, \mathcal{G}_\beta)$, based in Eq. (4), the overall energy of the system becomes a probabilistic sum over the set of energies obtained from all the associated set of graphs:

$$\begin{aligned} \Delta E_{\{\mathcal{G}_\beta\}}^{\text{target} \leftarrow \text{Ref}}(\bar{\mathbf{x}}) &= \sum_{\beta} \sum_{r=0}^{\mathcal{R}_\beta} \sum_{\alpha_r^\beta \in \mathbf{V}_r^\beta} \rho_\beta(\bar{\mathbf{x}}) \mathcal{M}_{\alpha_r^\beta, r}^{\mathcal{R}_\beta} \Delta E_{\alpha_r^\beta, r}(\bar{\mathbf{x}}) \\ &= \sum_{\beta} \sum_{r=0}^{\mathcal{R}_\beta} \sum_{\alpha_r^\beta \in \mathbf{V}_r^\beta} \tilde{\rho}_\beta(\bar{\mathbf{x}}) \Delta E_{\alpha_r^\beta, r}(\bar{\mathbf{x}}). \end{aligned} \quad (6)$$

The graph weights $\tilde{\rho}_\beta(\bar{\mathbf{x}})$ are obtained variationally in Refs. [43,47] and through numerical considerations in Ref. [48].

In this publication, we consider only a single graph topology, and we show that a family of neural networks used to train the quantities $\{\Delta E_{\alpha_r,r}(\bar{\mathbf{x}}_{\alpha_r,r}) \rightarrow \Delta E_{\alpha_r,r}^{\text{ML},1}(\bar{\mathbf{x}}_{\alpha_r,r})\}$ and obtain an accurate potential energy surface for one system, that is, $\{\Delta E_{\alpha_r,r}^{\text{ML},1}(\bar{\mathbf{x}}_{\alpha_r,r})\} \rightarrow E_{\text{system},1}^{\text{target}}(\bar{\mathbf{x}})$,

$$\begin{aligned} \Delta E_{\text{system},1}^{\text{target} \leftarrow \text{Ref}}(\bar{\mathbf{x}}) &= E_{\text{system},1}^{\text{target}}(\bar{\mathbf{x}}) - E_{\text{system},1}^{\text{Ref}}(\bar{\mathbf{x}}) \\ &= \sum_{r=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r,r}^{\mathcal{R}} \Delta E_{\alpha_r,r}^{\text{ML},1}(\bar{\mathbf{x}}_{\alpha_r,r}), \end{aligned} \quad (7)$$

can be refined in a well-defined and consistent numerical manner to yield an accurate potential for a much larger system, that is,

$$\Delta E_{\alpha_r,r}^{\text{ML},1}(\bar{\mathbf{x}}_{\alpha_r,r}) \xrightarrow{\text{refine}} \Delta E_{\alpha_r,r}^{\text{ML},2}(\bar{\mathbf{x}}_{\alpha_r,r}) \quad (8)$$

and $\{\Delta E_{\alpha_r,r}^{\text{ML},2}(\bar{\mathbf{x}}_{\alpha_r,r})\} \rightarrow E_{\text{system},2}^{\text{target}}(\bar{\mathbf{x}})$, yielding

$$\begin{aligned} \Delta E_{\text{system},2}^{\text{target} \leftarrow \text{Ref}}(\bar{\mathbf{x}}) &= E_{\text{system},2}^{\text{target}}(\bar{\mathbf{x}}) - E_{\text{system},2}^{\text{Ref}}(\bar{\mathbf{x}}) \\ &= \sum_{r=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r,r}^{\mathcal{R}} \Delta E_{\alpha_r,r}^{\text{ML},2}(\bar{\mathbf{x}}_{\alpha_r,r}). \end{aligned} \quad (9)$$

Thus, the key idea in this paper is captured in Eq. (8). In this publication, specifically, we refine the graph-theory-based fragment neural networks used to construct the potential energy surface for solvated Zundel to then obtain the surface for the protonated 21-water cluster. In this way, we provide an algorithm for growth in learning, and our approach looks very similar to the transformer architecture used in large language models [37]. This connection is discussed in Sec. III. To explicitly outline how we build our family of neural network models with reduced network depth as well as breadth, we outline a network projection technique in the section below to compute machine-learning (ML) approximations to Eq. (2).

A. Neural network projections from the graph-theoretic scheme

To arrive at Eq. (2), we begin with a general graph-based Hilbert space decomposition scheme and illustrate the procedure using Fig. 4. To illustrate the idea, we begin with a general partition of R^n (the n -dimensional real coordinate space). (We may also consider C^n , but for simplicity here we partition R^n .) The electronic Hilbert space that defines the full molecular system may, for simplicity, be represented using the coordinate representation $\{|x\rangle\}$, and these elements are members of R^n . Hence, we may then divide the molecular space into regions such as A , B , etc., as shown in Fig. 4. At this stage, the boundaries of these regions are not considered to be strict, and we may choose these regions based on some prior chemical insight. For example, the electrons are confined

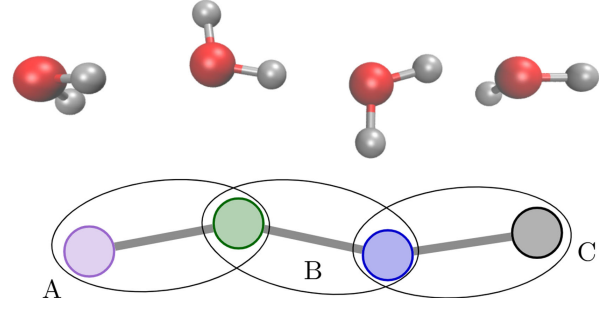


FIG. 4. Illustration of sets and graphs described in Sec. II A. The water wire on top is represented as a graph, with nodes being individual water molecules, and the graph also defines the sets A , B , and C . See Fig. 5 for a graphical representation of systems studied in this paper.

by attraction to the nuclear framework, and one may use this idea to divide the molecular space into chemically meaningful regions. Examples of how such a partitioning can be carried out can be found in Appendix A, for both bonded as well as nonbonded systems. We may then resolve the basis set resolution of identity, using individual projectors, $\{\mathcal{P}_A, \mathcal{P}_B, \dots\}$, along with the set-theoretic inclusion-exclusion principle [75] as

$$\begin{aligned} \mathbf{I} &= \mathcal{P}_A + \mathcal{P}_B + \mathcal{P}_C + \dots \\ &\quad - \mathcal{P}_{A \cap B} - \mathcal{P}_{A \cap C} - \mathcal{P}_{B \cap C} - \dots \\ &\quad + \mathcal{P}_{A \cap B \cap C} + \dots \\ &\quad \dots, \end{aligned} \quad (10)$$

where

$$\mathcal{P}_A = \int_A dx |x\rangle \langle x|. \quad (11)$$

Thus, the regions A , B , etc., are partitions of a molecular Hilbert space, and for the coordinate representation, the spatial decomposition may be seen from Fig. 4. In Refs. [49,50,53,54], we have shown that there is another, more general and robust approach to construct such a resolution using a graph-theoretic representation of molecular fragmentation, which divides the molecular electronic

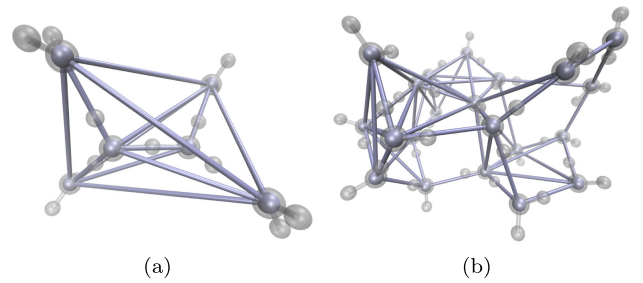


FIG. 5. The graphical complexity for the two systems.

Hilbert space using nodes, edges, faces, and higher-rank simplexes [76–78] that constitute a graph. This yields an alternative way to represent the molecule’s electronic Hilbert space through a family of projection operators $\mathcal{P}_{\alpha_r,r}$ to obtain [49,50,53,54]

$$\mathbf{I} = \sum_r \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r,r} \mathcal{P}_{\alpha_r,r}, \quad (12)$$

where

$$\mathcal{P}_{\alpha_r,r} = \int_{\forall x \in \alpha_r} dx |x\rangle \langle x|. \quad (13)$$

The notation α_r , r , and $\mathbf{V}_r$ are as described at the top of Sec. II A. When the operator in Eq. (12) is applied to a molecular system, it will decompose the system into fragments represented using nodes (local, connected, or interacting groups of atoms), edges (node-node interactions, and this may include connected, bonded [39,41,42,45,46], or disconnected, and, hence, through space, nonbonded interactions [38–40,43–45,47–51]), faces (three-node interactions), tetrahedrons (four-node multibody interactions), etc., and produce an energy expression [38–51]:

$$E_{\text{graph}}^{\text{target}}(\bar{\mathbf{x}}) = \sum_{r=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r,r}^{\mathcal{R}} E_{\alpha_r,r}^{\text{target}}(\bar{\mathbf{x}}_{\alpha_r,r}), \quad (14)$$

and target here is some target level of electronic structure theory, such as CCSD, as before. The application of Eq. (12) onto the full system, hence, yields a set of fragment energies $\{E_{\alpha_r,r}^{\text{target}}(\bar{\mathbf{x}}_{\alpha_r,r})\}$ that then provide an approximation to the total energy as per Eq. (14). We also note that Eq. (12) is simply an inclusion-exclusion principle [75] interpretation of the resolution of identity within the graph structure and

$$\begin{aligned} \mathbf{I} &= \sum_r \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r,r} \int_{\forall \bar{\mathbf{x}}_{\alpha_r,r} \in \alpha_r} d\bar{\mathbf{x}}_{\alpha_r,r} |\bar{\mathbf{x}}_{\alpha_r,r}\rangle \langle \bar{\mathbf{x}}_{\alpha_r,r}| \\ &= \sum_r \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r,r} \mathcal{P}_{\alpha_r,r}. \end{aligned} \quad (15)$$

Connections between Eq. (14) and the commonly used many-body expansions have been discussed in Refs. [41,42,48]. For this to be the case, it is required that the graph above be a simplicial complex [76–84]. However, it has been shown in Refs. [44,48] that such an approximation converges slower with the maximum truncation order $\mathcal{R}$ above, as compared to Eqs. (2) and (4), where this idea is included within an our own N-layered integrated molecular orbital and molecular mechanics (ONIOM)-like [85] formalism and energy differences are used.

However, as the complexity and size of the system grow, we find that this approach also becomes expensive by needing post-Hartree-Fock energies and gradients for larger and larger clusters obtained from higher-rank objects discussed above. That is, $\{E_{\alpha_r,r}^{\text{target}}(\bar{\mathbf{x}}_{\alpha_r,r})\}$ may be needed for larger values of r . Furthermore, the number of fragments produced would also grow rapidly (in a combinatorial fashion) as the system size grows, since higher-rank objects can be formed by all possible combinations of nodes within a distance cutoff. In Refs. [46,49], we provided machine-learning approximations for the $\Delta E_{\alpha_r,r}$ values in larger clusters, to scale down the cost of obtaining potential energy surfaces for large gas-phase and condensed-phase problems. In other words, we use neural network predictions for $\Delta E_{\alpha_r,r}^{\text{ML}}(\bar{\mathbf{x}}_{\alpha_r,r})$ to replace the larger-rank $\Delta E_{\alpha_r,r}(\bar{\mathbf{x}}_{\alpha_r,r})$ values in Eq. (3), as shown in Eqs. (7) and (9). Thus, the approach here and in Refs. [46,49] may also be seen as being related to the Δ – ML ideas in the literature [86–90] but is, in fact, couched within the ONIOM formalism as discussed in Ref. [38]. As seen from Eqs. (7) and (9), the ideas here exploit the graph-based fragmentation technique that helps partition the learning problem. This is achieved by first assuming the existence of some neural network learning model $\mathcal{Q}$, which provides the energy of the system at the target post-Hartree-Fock level:

$$\bar{\mathbf{x}} \xrightarrow{\mathcal{Q}} [E^{\text{target}}(\bar{\mathbf{x}}) - E^{\text{Ref}}(\bar{\mathbf{x}})] \equiv \Delta E_{\text{system}}^{\text{target} \leftarrow \text{Ref}}(\bar{\mathbf{x}}). \quad (16)$$

Such a neural network $\mathcal{Q}$ may be visually represented, for example, by the network in Fig. 1. We may then simplify $\mathcal{Q}$ using Eq. (12) to obtain

$$\begin{aligned} \mathbf{I} \mathcal{Q} &\approx \left[\sum_{r=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r,r} \mathcal{P}_{\alpha_r,r} \right] \mathcal{Q} \\ &= \sum_{r=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r,r} \mathcal{Q}_{\alpha_r,r}. \end{aligned} \quad (17)$$

The full system neural network mapping $\mathcal{Q}$ now has been decomposed into a family of much smaller neural networks $\{\mathcal{Q}_{\alpha_r,r}\}$ that depends only on the corresponding fragment geometry. Consequently, the cost of preparing training samples from fragments becomes much easier, and the complexity of neural networks is also largely reduced [46,49,54]. The corresponding energy expression in Eq. (2) then becomes a delta machine-learning formalism [86–90], as the learning algorithm in Eq. (17) is modified to

$$E_{\text{ML}}^{\text{target}}(\bar{\mathbf{x}}) - E^{\text{Ref}}(\bar{\mathbf{x}}) = \sum_{r=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r,r}^{\mathcal{R}} \Delta E_{\alpha_r,r}^{\text{ML}}(\bar{\mathbf{x}}_{\alpha_r,r}), \quad (18)$$

and the ML error is

$$\begin{aligned}
E_{\text{ML}}^{\text{target}}(\bar{\mathbf{x}}) - E^{\text{target}}(\bar{\mathbf{x}}) &= \sum_{r=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r, r}^{\mathcal{R}} (\Delta E_{\alpha_r, r}^{\text{ML}}(\bar{\mathbf{x}}_{\alpha_r, r}) - \Delta E_{\alpha_r, r}(\bar{\mathbf{x}}_{\alpha_r, r})) \\
&\leq \sum_{r=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_r} |\mathcal{M}_{\alpha_r, r}^{\mathcal{R}}| |\Delta E_{\alpha_r, r}^{\text{ML}}(\bar{\mathbf{x}}_{\alpha_r, r}) - \Delta E_{\alpha_r, r}(\bar{\mathbf{x}}_{\alpha_r, r})|, \quad (19)
\end{aligned}$$

where we have used the Cauchy inequality [91] to provide an upper bound for the error expression.

The simplification afforded by this approximation may be seen from the visuals provided in Figs. 1–3, where we have shown how the neural networks for $\{\Delta E_{\alpha_r, r}^{\text{ML}}(\bar{\mathbf{x}}_{\alpha_r, r})\}$ can be used to construct approximations to $E_{\text{ML}}^{\text{target}}(\bar{\mathbf{x}})$. In Fig. 2, we show the independent neural networks for each node that contribute to $\{\Delta E_{\alpha_r, r=0}^{\text{ML}}(\bar{\mathbf{x}}_{\alpha_r, r=0})\}$ terms in Eq. (18). Similarly, in Fig. 3, we show the independent neural networks for each edge, and these two figures visually convey the reduced complexity of obtaining these fragment neural networks as can be seen by comparison with Fig. 1. In previous work [49], we have successfully constructed the neural network potential energy surface with kJ/mol mean absolute error for the solvated Zundel following Eq. (18) by using only 10% of fragments. In this paper, as seen in Eq. (9), we ask if the graph-theoretically decomposed neural networks used to obtain the potential energy surface for system 1 can be refined to obtain the same for system 2. Additionally, in the next section, we discuss how the projection operators discussed above connect to the idea of attention in natural language processing in modern large language models.

III. THE GRAPH ABSTRACTION CONNECTS MOLECULAR POTENTIAL ENERGY SURFACES AND NATURAL LANGUAGE PROCESSING

Current methodologies in natural language processing involve mapping linguistic components such as words or phrases in a sentence to abstract vector spaces. The semantic relations between these words are then represented as distances between these vector representations of the primary linguistic components (which, as stated above, could be words or phrases). This technique is used in large generative AI models. In Fig. 6, we show how such an abstraction of natural language can be related to our graph-theoretic depiction of molecular structure discussed in the previous section. The key similarity between both ideas is that the neighborhood of a word or phrase defines its context and semantics in natural language, much like the neighborhood of a molecular fragment, as captured from the graphical representation above, defines its stability. In the graph-theoretic formalism, the neighborhoods are defined using graph components (nodes, edges, faces, etc., as described within the sets $\mathbf{V}_r$). Each component of a molecular system, such as a fragment, has many

Mathematical components	Instance in Molecular potential energy surfaces	Instance in a language
Graph: Nodes	Groups of atoms or molecular fragments	Words or phrases (token) in a sentence.
Graph: Edges	Interaction between molecular fragments	Concatenated words or phrases. Sentence Grammar and semantics.
Graph: Faces	Higher order interaction between fragments	Larger language components from concatenated set of words or phrases.

FIG. 6. The common constructs of the graph-based depiction of molecular potential energy surfaces used here and elements of a LLM.

different possible energy values arising from multiple possible configurations, and the stability of each of these configurations depends on how the fragment adapts to its surroundings (as dictated by the edges, faces, and higher-order simplexes in a graph). Likewise, parts of a sentence, such as a word or a phrase, may carry different meanings with some probabilistic distribution, and the intended meaning arises given the presence of other language components (such as surrounding words and phrases) and provides a specific context. In our work, the weight of each molecular fragment has a critical role that arises completely from geometric considerations as enforced by graphical constraints and allows us to compute the full system potential energy as a weighted sum of fragment energies as given by Eq. (2).

In a complementary fashion, in natural language processing, the attention mechanism within transformers [37] determines the importance and contextual meaning of each word by analyzing the relationship between all possible pairs of words within a sentence or a paragraph. In the transformer, each word or token is represented by an embedding vector. This embedding vector is then linearly transformed to three vectors: *query*, *key*, and *value*. The query vectors encode querying information about a specific word, whereas the key vectors contain potential maps to other known words. Thus, for the i th word in a sentence, there are three associated vectors to compute the attention for the sentence, namely, Q_i for the query, K_i for the key, V_i for the value. The dot product $[Q_i \cdot K_j]$, that is, the dot product of the query vector for the i th word and the key vector for the j th word in the sentence, produces the relevance between the two words. Collecting all possible such inner products provides a cumulative relevance of a certain query vector $[Q_i K^T]$, where $K^T \equiv \{K_1, K_2, \dots\}$ contains all key vectors. By extension, the quantity $[Q K^T]$ simply projects all query vectors onto the key vector space. One critical idea introduced in Ref. [37], which is purportedly responsible [37,92] for the major generative AI revolution that we are witnessing at the moment, is to scale $[Q K^T]$ using a softmax function given by

$$\left[\text{softmax} \left(\frac{Q K^T}{\sqrt{d}} \right) \right]_{i,j} = \frac{\exp\{Q_i \cdot K_j / \sqrt{d}\}}{\sum_k \exp\{Q_i \cdot K_k / \sqrt{d}\}}, \quad (20)$$

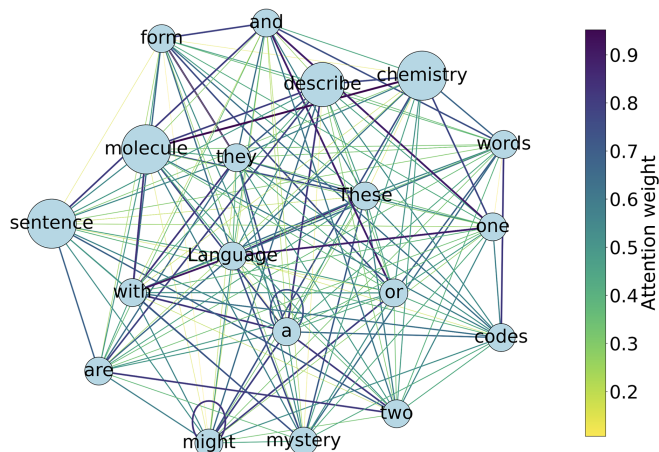


FIG. 7. An attention mechanism network developed for the sentence: “These words might form a sentence, or they might describe a molecule. Language and chemistry are two codes with one mystery.” This network is generated using an AI agent in Microsoft Copilot [115]. The various steps involved in this process are as follows. Tokenization: the sentence was split into 15 lowercase tokens. Embedding simulation: each token was assigned a random eight-dimensional vector (simulating word embeddings). Query, key, and value matrices: these were generated using random weight matrices in this case, for illustrative purposes. Attention scores: computed using the dot product of query and key matrices, scaled by the square root of the embedding dimension. Softmax: applied to get attention weights. Attention output: calculated as the weighted sum of the value vectors. Also see Fig. 8, where these algorithmic aspects are compared with those in the current molecular scheme.

where d is the dimensionality of each query vector and key vector. Thus, the softmax function, in Boltzmann style, simply weights the most significant key vectors that contribute to all queries. An illustration of this softmax function is provided in Fig. 7. Here, we have presented the sentence: “These words might form a sentence, or they might describe a molecule. Language and chemistry are two codes with one mystery.” The nodes in the graph are the words from the sentence, and the colors of the edges represent the softmax function value for the words connected by the respective edges. The way in which this graph is generated is described in the caption of Fig. 7. For example, in this sentence, the network has allowed us to discover that the words “molecule” and “language” have one of the highest softmax function values. But this network was developed based on only one sentence, and clearly the networks developed using more data can provide us with far greater information on word correlations.

The next important part of the transformer architecture [37] is the value vector space. The *key-value* pairs together transform the queries to contextual information. Thus [37],

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (21)$$

or, more explicitly,

$$[\text{Attention}(Q, K, V)]_{i,l} = \sum_j \frac{\exp\{Q_i \cdot K_j / \sqrt{d}\}}{\sum_k \exp\{Q_i \cdot K_k / \sqrt{d}\}} V_{j,l}, \quad (22)$$

where the elements $\{V_{j,l}\}$ provide the contextual meaning of the j th key vector. As a result, an attention network like that created in Fig. 7 allows a query $\rightarrow$ value map based on the context within a sentence and helps decipher the meaning of a sentence.

In the same manner, our graph-theoretic approach in Eq. (2) divides a molecule into *simplexes* or rank- r objects, and the weight of each such rank- r object is given by $\mathcal{M}_{\alpha,r}^R$. These quantities act as weights on the $\Delta E_{\alpha,r}(\bar{\mathbf{x}}_{\alpha,r})$ values.

Based on the above discussion, we may make the connection that the softmax function has a similar role as $\mathcal{M}_{\alpha,r}^R$ in Eq. (2), where $\mathcal{M}_{\alpha,r}^R$ also provides the importance of the α th, r -rank fragment (or simplex). Hence, we may state that

$$\text{Importance} \equiv \left[\text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right) \right]_{i,j} \leftrightarrow \mathcal{M}_{\alpha,r}^R. \quad (23)$$

Importantly, both terms in Eq. (23) arise from purely geometric considerations. Geometry in the context of the molecular system is the spatial spread of fragments, whereas geometry in the context of natural language processing is connectivity or conjunction of words or phrases.

Thus, it appears that Eqs. (22) and (2) carry out importance-to-value projections where

$$\text{intrinsic value} \equiv V_{j,l} \leftrightarrow \Delta E_{\alpha,r}(\bar{\mathbf{x}}). \quad (24)$$

As a comparison to drive home the message, we also provide Fig. 8, which may be compared with Fig. 7.

This connection can be made even more explicit by rewriting attention in bra-ket notation. In the absence of the softmax function, this map essentially writes the query vectors in the biorthogonal $|K\rangle\langle V|$ basis and, thus, in the absence of softmax,

$$[\text{Attention}(Q, K, V)]_{\omega,v} = \sum_{\kappa} \langle Q_{\omega} || K_{\kappa} \rangle \langle V_v | \quad (25)$$

and in the presence of softmax

$$[\text{Attention}(Q, K, V)]_{\omega,v} = \sum_{\kappa} f(\langle Q_{\omega} || K_{\kappa} \rangle) \langle V_v |. \quad (26)$$

Thus, the projections, $\langle Q_{\omega} || K_{\kappa} \rangle$ or $f(\langle Q_{\omega} || K_{\kappa} \rangle)$, act in an analogous fashion as the projections $\mathcal{P}_{\alpha,r}$ in Eq. (12).

To complement the discussion above, we present the various stages of LLM generation and graph-theoretic decomposition in Fig. 9 to make the connections between

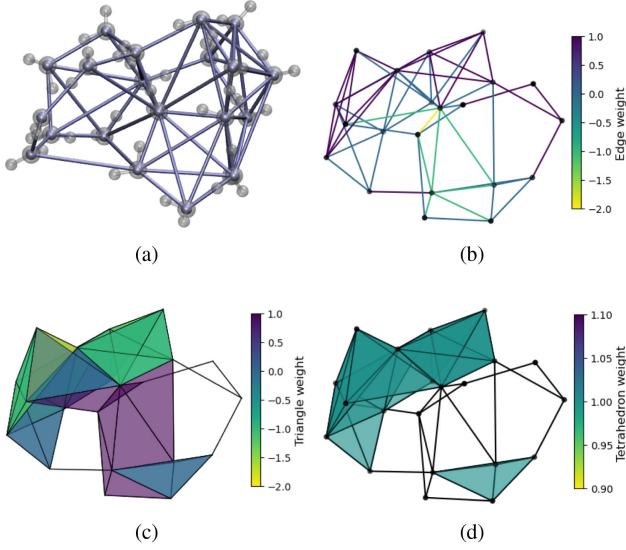


FIG. 8. For comparison with Fig. 7. (a) Graph rendering of the protonated 21-water cluster. (b) For the molecule in (a), edge colors now represent $\mathcal{M}_{\alpha,r}^R$ values. Compare with Fig. 7. Thus, these edges weights can be negative here, unlike in Eq. (20). Furthermore, Eq. (2) allows higher-rank objects, too (and not just edges). Similar to (b), the graph-theoretic $\mathcal{M}_{\alpha,r}^R$ values exists for all higher-ranked objects, such as triangles in (c) and tetrahedrons in (d), with $\mathcal{M}_{\alpha,r}^R$ weights presented here. The white space in the graph is due to no triangle or tetrahedron formed at those positions, as the nodes are not all mutually connected. The overall molecular energy (or stability) is given by the weighted sum of simplex energy values $\Delta E_{\alpha,r}(\bar{\mathbf{x}}_{\alpha,r})$ with corresponding simplex weights $\mathcal{M}_{\alpha,r}^R$. Compare Eqs. (2) and (22).

processes involved more explicit. A few distinctions between challenges that determine the complexity of chemical potential energy surfaces and creation of LLMs are discussed in Appendix D.

As an illustration of our approach, we show how simplex neural networks that were originally defined for obtaining accurate potential energy surfaces for the solvated Zundel system can now be generalized to obtain results for the much larger protonated 21-water cluster. Thus, this paper provides the first sub-kcal/mol-level neural network description of larger protonated water clusters.

Algorithmic steps	LLM analogue	Chemical analogue
Tokenization	Split sentence of paragraphs into words or phrases.	Molecule is represented as a graph which then leads to molecular fragments
Embedding Simulation	Develop <i>Query</i> , <i>Key</i> , and <i>Value</i> Matrices for all tokens	Obtain (α, r) ; $\Delta E_{\alpha,r}$ If needed develop NN approximations for $\Delta E_{\alpha,r}$
Attention Weights	Compute the <i>softmax</i> function values for all edges.	Compute $M_{\alpha,r}^R$

FIG. 9. The common algorithmic aspects involved in the natural language process and the graph-theoretic process presented here.

IV. POTENTIAL ENERGY SURFACES FOR SOLVATED ZUNDEL USING THE FAMILY OF NEURAL NETWORKS INTRODUCED IN SEC. II

The solvated Zundel cation has a fundamental role in the study of proton transfer in water, as these contain two of the most critical substructures, the Eigen and Zundel moieties of water [93–97]. The interconversion between the Eigen and Zundel systems offers valuable insight into proton transfer in a wide range of biological, atmospheric, material, and condensed phase systems [48,66,98–108]. However, due to the highly polarized nature of the Eigen and Zundel cations, accurately modeling these systems typically requires computationally prohibitive post-Hartree-Fock calculations [109–113] along with detailed description of multidimensional quantum nuclear effects. This makes the construction of accurate potential energy surfaces and the associated quantum dynamical study extremely challenging [62,63,66]. In this paper, we choose the solvated Zundel as a foundational test case to (a) develop neural networks trained on molecular fragments of solvated Zundel that are then combined as per Eq. (7) to reliably reproduce the full system potential energies of solvated Zundel, and (b) since the solvated Zundel system serves as a fundamental model for studying proton migration, we then probe whether the successful recovery of its potential energy surface provides a reliable starting point to use the same family of fragment neural network models to study more complicated and larger systems, such as the protonated 21-water clusters.

To construct a set of fragment neural networks for the solvated Zundel system, we follow a procedure similar to that in Ref. [49]. Briefly, we begin by preparing an AIMD trajectory for the solvated Zundel system, by applying graph-theoretic fragmentation to obtain fragments from each structure, and create a data bank for each fragment type. These AIMD geometries use the energy expressions in Eq. (2) along with gradients discussed in Refs. [38–48,51]. The quantity E^{Ref} is chosen as DFT with 6-31++g(d,p) basis, and the E^{target} level is CCSD with 6-31++g(d,p) basis [40]. We considered at most rank-3 simplexes [$\mathcal{R} = 3$ in Eq. (2) in these AIMD trajectories and, hence, four-body interactions between the nodes] with maximum oxygen-oxygen distance (which defines the maximum edge length to define the graph) values set at 7.5 Å [48]. Additionally, the hydrogen-oxygen distance cutoff value is set to be 1.4 Å and defines the chemical bonding environment used to obtain the graphs and fragments. These choices are based on our previous studies [38–40,48] and help determine the graph and associated fragments for any given structure. Our library is an in-house PYTHON code that uses NUMPY and SCIPY functionalities. The numbers of fragments of each kind obtained during these trajectories are summarized in Table I.

We then apply a minibatch k -means clustering algorithm, which was discussed in Ref. [49] and is summarized further

TABLE I. Number of fragments in the solvated Zundel library, the corresponding number of data points in a training library, and the library size for the protonated 21-water cluster system used for testing predictions.

Fragment	$(\text{H}_2\text{O})_6\text{H}^+$ data ^a		$(\text{H}_2\text{O})_{21}\text{H}^+$ data ^b
	Full data	Training dataset size ^c	Target data
H_2O	38 910	3891	559 877
H_3O^+	17 046	1704	34 297
H_4O_2	62 380	6238	1 356 707
H_5O_2^+	77 510	7751	190 727
$\text{H}_6\text{O}_2^{++\text{d}}$	0	0	99
H_6O_3	46 940	4694	1 014 960
H_7O_3^+	139 580	13 958	257 260
$\text{H}_8\text{O}_3^{++\text{d}}$	0	0	256
H_8O_4	15 750	1575	253 889
H_9O_4^+	124 140	12 414	108 123
$\text{H}_{10}\text{O}_4^{++\text{d}}$	0	0	71

^aNumber of fragments from the $(\text{H}_2\text{O})_6\text{H}^+$ trajectory.

^bNumber of fragments from the $(\text{H}_2\text{O})_{21}\text{H}^+$ trajectory.

^cNeural network training data size. These data points are obtained from a multidimensional k -means clustering algorithm [49,114] of the full $(\text{H}_2\text{O})_6\text{H}^+$ data and represent 10% of the dataset.

^dVery low probability doubly protonated fragments found in the larger 21-water cluster but not in solvated Zundel.

below in this section and in more detail in Appendix B. This helps us select 10% of the data for each type of fragment to be used for training and construct separate neural network models to obtain the corresponding fragment energy surfaces. Since there are multiple protons present inside each fragment, we include permutation symmetry into our treatment for all fragments, as explained in Appendix C. We have also adopted Gaussian functions as the activation functions to improve the smoothness and differentiability of the model predictions. At the end of this training cycle, we determine the fragment energy surfaces, $\Delta E_{\alpha,r}^{\text{ML},1}(\bar{\mathbf{x}}_{\alpha,r})$ needed in Eq. (7). This helps develop a model of the full potential energy surface as indicated on the left side of Eq. (2).

In Fig. 10, we demonstrate the accuracy of this ML architecture for the full system potential energy surface of the solvated Zundel problem [left side of Eq. (7)]. The axis labels in Fig. 10 are shifted relative to the lowest energy configuration found in the solvated Zundel trajectory. The horizontal axis represents the ML result [right side of Eq. (7)], and the vertical axis represents the full CCSD direct calculation [$\mathcal{O}(N^6)$ scaling of cost for vertical axes]. The error (spread of the distribution away from the solid diagonal line) distribution is also shown in Fig. 10. Clearly, the neural networks have been well trained for solvated Zundel with 10% training at rank 3. We arrive at a full system mean absolute error (MAE) of only 0.36 kcal/mol for the potential energies for the solvated Zundel trajectory.

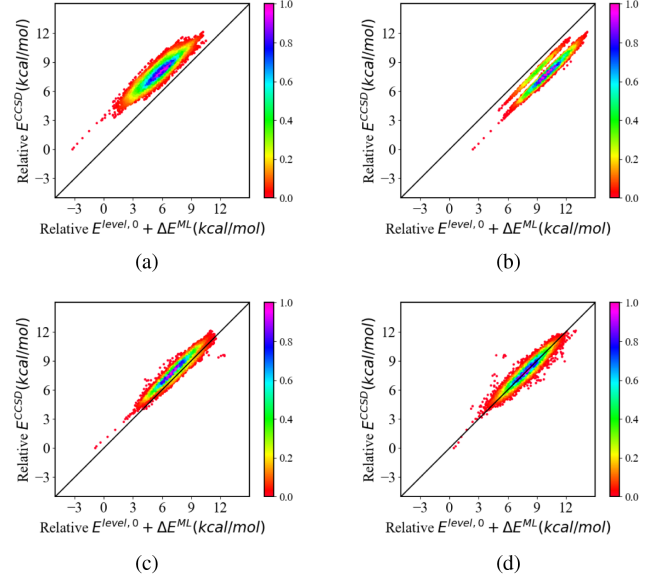


FIG. 10. The solvated Zundel potential energy absolute errors. All energies are plotted relative to the minimum potential energy of solvated Zundel in the trajectory. (a) Using graph-based NN models for solvated-Zundel, $\mathcal{R} = 0$. (b) Using graph-based NN models for solvated-Zundel, $\mathcal{R} = 1$. (c) Using graph-based NN models for solvated-Zundel, $\mathcal{R} = 2$. (d) Using graph-based NN models for solvated-Zundel, $\mathcal{R} = 3$.

As noted above, the $\Delta E_{\alpha,r}^{\text{ML},1}(\bar{\mathbf{x}}_{\alpha,r})$ needed to compute the left side in Eq. (7) are obtained through a minibatch k -means procedure. We begin with the training a set of neural networks following Appendix B for the fragments in Table I using the solvated Zundel data. To develop these neural networks, we generate training sets using the k -means clustering [49,114] method. We perform k -means clustering [114] calculations on the database of fragment structures obtained from solvated Zundel AIMD trajectory to create a representative training set for the fragments in the solvated Zundel system. Using k means, we divide or tessellate the data space for each fragment type in Table I into regions, and these regions are called clusters. Each cluster is represented by a centroid that is computed as the arithmetic mean of all data points in that specific cluster. This primitive training set (that is, the set of centroids obtained from k means) is chosen to be 10% in size (as compared to the full dataset for the solvated Zundel system in Table I) and is used to construct neural networks and fragment energies needed in Eq. (18). Examples of such neural networks are shown in Figs. 2 and 3. These neural networks are used to then compute energy values for the data from the full trajectory for solvated Zundel as well as the 21-mer with fragment energy errors for the target level of the theory given in Table II and the distribution of errors for full system geometries for both systems being given in Fig. 11.

TABLE II. MAE in kcal/mol from initial models trained on 10% solvated Zundel fragments (see Table I) and predictions for both solvated Zundel fragments and protonated 21-water cluster fragments.

Fragments	$(\text{H}_2\text{O})_6\text{H}^+$		$(\text{H}_2\text{O})_{21}\text{H}^+$	
	$\omega_{f_r, \mathcal{R}}^a$	MAE ^b	$\omega_{f_r, \mathcal{R}}^c$	MAE ^d
H_2O	0.13	0.00	0.14	0.00
H_3O^+	0.06	0.00	0.01	0.00
H_4O_2	0.16	0.01	0.30	0.06
H_5O_2^+	0.20	0.01	0.04	0.15
H_6O_3	0.08	0.01	0.28	0.31
H_7O_3^+	0.24	0.02	0.06	0.45
H_8O_4	0.01	0.04	0.12	0.59
H_9O_4^+	0.11	0.05	0.05	0.88

^aFor fragments found in the $(\text{H}_2\text{O})_6\text{H}^+$ AIMD trajectory at $R = 3$ from Eq. (14).

^bMAE for fragments found in the $(\text{H}_2\text{O})_6\text{H}^+$ AIMD trajectory.

^cFor fragments found in the $(\text{H}_2\text{O})_{21}\text{H}^+$ AIMD trajectory at $R = 3$ from Eq. (14).

^dMAE for fragments found in the $(\text{H}_2\text{O})_{21}\text{H}^+$ AIMD trajectory.

A. The challenge of transferring fragment neural networks across molecular systems

Next, we assess the quality of predictions from these fragment neural network models in obtaining the protonated 21-water clusters. Figure 5 is used to show the level of complexity of the graphical representation as one moves from the solvated Zundel system to the protonated 21-water cluster. Specifically, assuming $O(N^6)$ complexity for the CCSD electronic structure values used here, each electronic structure energy for the system in Fig. 5(b) is 1762 times more expensive to compute than for that in Fig. 5(a). When one accounts for the fact that the system in Fig. 5(b) has 186 independent degrees of freedom, whereas the system in Fig. 5(a) contains *only* 51 degrees of freedom, the prohibitive exponential complexity associated with training for the potential energy surface for the system in Fig. 5(b) is apparent.

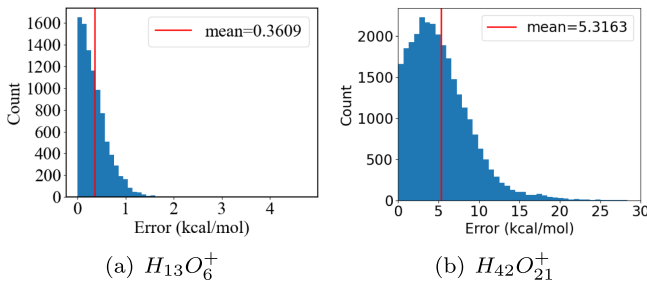


FIG. 11. (a) Complements Fig. 10, and (b) complements Fig. 12. The absolute error distribution (blue) and MAE (red) for solvated Zundel (a) and protonated 21-water cluster (b) full system potential energies.

As before, to gauge the accuracy of the solvated Zundel neural network fragments in being able to produce accurate protonated 21-water cluster energies, we first develop a library of protonated 21-water cluster structures. Thus, in addition to the solvated Zundel AIMD trajectory noted earlier, we also generate a new trajectory for the protonated 21-water cluster including 28 294 frames. The corresponding library of structures include 9326 geometry frames from the solvated Zundel described in the previous section and 28 294 frames from the protonated 21-water cluster trajectory. These AIMD geometries (as for the solvated Zundel trajectories previous noted) are obtained using the graph-theoretic fragmentation energy expressions [Eq. (2)] along with gradients as discussed in Refs. [38–48,51]. The levels of theory, basis, maximum graph rank are the same as in the case of solvated Zundel. The maximum edge length to define the graph is set at 7.5 Å for the solvated Zundel and 4.5 Å for the protonated 21-water cluster, respectively [48]. Additionally, as before, the hydrogen-oxygen distance cutoff value is set to be 1.4 Å and defines the chemical bonding environment used to obtain the graphs and fragments. These choices are based on our previous studies [38–40,48]. The number of fragments of each kind obtained during both trajectories is summarized in Table I.

However, when the fragment neural network models trained from data obtained in the solvated Zundel AIMD trajectory are used to compute the potential energy surface for the protonated 21-water cluster, an increase in error is noted for larger values of rank $\mathcal{R}$, as seen in Figs. 12(a)–12(d). Also see Fig. 11. Unlike the solvated Zundel, the full system CCSD energy calculations are prohibitive. Hence, the vertical axis in Figs. 12(a)–12(d) are E^{target} from Eq. (2), and the error is essentially the error in ML predictions as given by Eq. (19). As can be seen from Figs. 12(a)–12(d), when the primitive solvated Zundel neural network models $[\Delta E_{\alpha_r, r}^{\text{ML}, 1}(\bar{\mathbf{x}}_{\alpha_r, r})]$ in Eq. (7) are applied, the errors get worse with increasing rank. We drill down into the reasons behind the errors in Figs. 12(a)–12(d) in Sec. IV A.

But there are two fundamental reasons for these differences.

First, as the rank increases, the number of fragments as well as the number of $\Delta E_{\alpha_r, r}^{\text{ML}, 1}(\bar{\mathbf{x}}_{\alpha_r, r})$ terms increases in Eq. (2). This can be seen from Fig. 13. Additionally, the error is bounded by Eq. (19), and the bound grows with the number of fragments of a certain size. In Table II, we also present the quantity relative significance of all rank- r fragments in Eq. (2). Toward this, we first introduce

$$\omega_{r, \mathcal{R}} = \frac{\sum_{\alpha_r \in \mathbf{V}_r} |\mathcal{M}_{\alpha_r, r}^{\mathcal{R}}|}{\sum_{r'=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_{r'}} |\mathcal{M}_{\alpha_r, r'}^{\mathcal{R}}|}, \quad (27)$$

which is the relative significance of all rank- r fragments in Eqs. (2) and (18). Additionally, a specific kind of fragment of rank- r , $f_r \subseteq \mathbf{V}_r$, may have significance given by

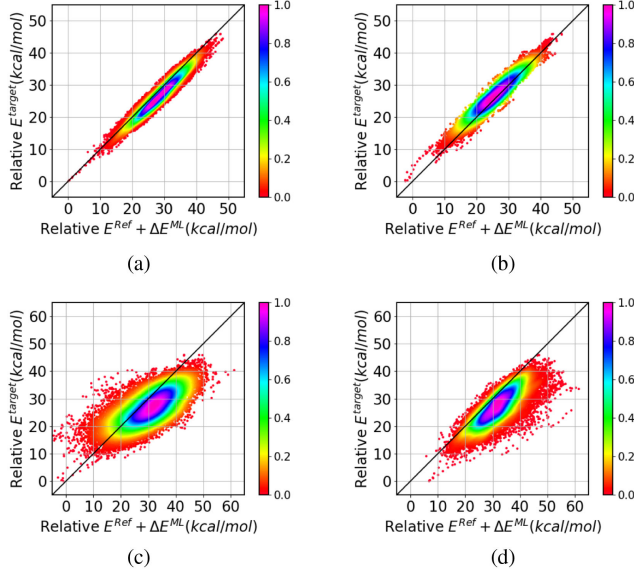


FIG. 12. Similar to Fig. 10, for the protonated 21-water cluster energies from NN models for the solvated Zundel system. Clearly, the error increases with increasing rank, since the potential energy surface space for the larger clusters in the 21-mer is not well represented in the original solvated Zundel data. (a) Using graph-based NN models for solvated-Zundel, $\mathcal{R} = 0$. (b) Using graph-based NN models for solvated-Zundel, $\mathcal{R} = 1$. (c) Using graph-based NN models for solvated-Zundel, $\mathcal{R} = 2$. (d) Using graph-based NN models for solvated-Zundel, $\mathcal{R} = 3$.

$$\omega_{f_r, \mathcal{R}} = \frac{\sum_{\alpha_r \in f_r} |\mathcal{M}_{\alpha_r, r}^{\mathcal{R}}|}{\sum_{r'=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_{r'}} |\mathcal{M}_{\alpha_r, r'}^{\mathcal{R}}|}. \quad (28)$$

This last quantity gauges the significance of each fragment in the overall target energy and is presented in Table II. We clearly see that some of the larger-sized fragments are more significant in the 21-water cluster than they were in the solvated Zundel system, which could be expected to have an important role on the overall error. Hence, the overall full system error accumulates and makes the quality of predictions unstable.

Second, higher-rank simplexes introduce larger fragments into the graph representation. These large fragments have more degrees of freedom and represent more complicated interactions within the molecular systems. More importantly, as we discuss later in Appendix E and in Sec. IV A, large fragments from the 21-mer occupy substantially different regions of the configuration space compared to those from the solvated Zundel. This difference arises from the greater flexibility for hydrogen transfer and geometry deformation in the larger cluster. Therefore, there is a need to adapt the solvated Zundel family of neural networks to an orthogonal to the original solvated Zundel data. We introduce such a scheme in the next section.

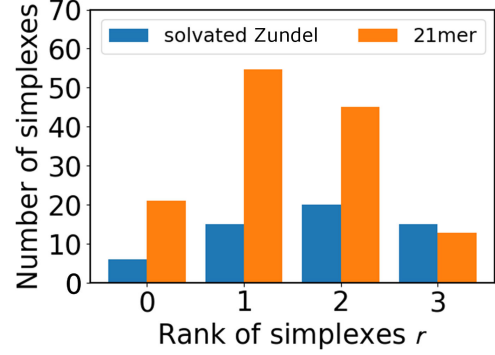


FIG. 13. The average number of simplexes for each rank obtained from the library of solvated Zundel and protonated 21-water cluster structures in the AIMD trajectories. The solvated Zundel trajectory library contains 9326 different geometries created using edge length cutoff 7.5 Å. The $r = 3$ fragment calculations are roughly $4^6 \equiv 4096$ times more expensive as compared to the $r = 0$ fragment calculations for CCSD accuracy. The protonated 21-water cluster (21-mer) AIMD library has 28 294 geometries but with edge cutoff 4.5 Å, so the average number of rank 3 simplexes is fewer. Simplexes with zero weights are not presented in the figure.

B. Incremental transformation of fragment network models

As can be seen from Table II and Fig. 11, for the 10% training, we can construct highly accurate models for solvated Zundel fragment energies and obtain full system potential energies with an MAE of 0.36 kcal/mol. Additionally, most of the structures are in the sub-kcal/mol range. However, when these same neural network models are used to evaluate the target energies for the structures in the larger protonated 21-water clusters, the full system error has raised to 5.32 kcal/mol, with a substantial population of structures with errors greater than 10 kcal/mol. This is because the larger fragments, beginning with the protonated water dimer systems, contain geometries that are very different in the protonated 21-water cluster trajectory as compared to the solvated Zundel trajectory. These differences in fragment structures are responsible for the larger errors in fragment energies in Table II and Fig. 11, and, as a consequence, the full system error accumulates rapidly with fragment errors according to Eq. (18). Our goal is to lay the ground work to transform, as needed, and transfer the neural networks computed from a solvated Zundel database to the larger protonated 21-water cluster database.

The range of fragment structures that significantly contribute to the solvated Zundel system could be very different from those that contribute to the protonated 21-water cluster system, and this may contribute to error trends noted in Fig. 12. To gauge the extent to which the range of structures differ between the two libraries, we first examine the differences in distributions of fragment geometries as they appear in a solvated Zundel system with that obtained for the

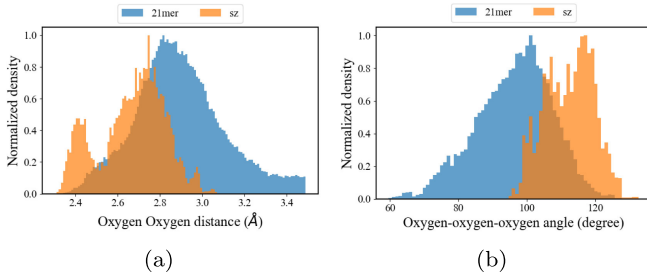


FIG. 14. Oxygen-oxygen distance distribution and oxygen-oxygen-oxygen angle distribution for the solvated Zundel and protonated 21-water cluster trajectories.

larger protonated 21-water cluster system. The larger system allows each fragment to, in general, sample a greater portion of the potential energy space. For example, this can be seen from the oxygen-oxygen (OO) radial distribution functions for the full systems as shown in Fig. 14(a) and the hydrogen bonding oxygen-oxygen-oxygen (OOO) angles as shown in Fig. 14(b). For the solvated Zundel trajectory, OO distances display two distinct peaks. The smaller one centers at 2.4 Å corresponding to the Zundel structures, and the other around 2.7 Å associates with the Eigen structures or the solvation water molecules surrounding the Zundel ions. By contrast, the protonated 21-water cluster displays a wider OO radial distribution. A similar trend is observed for the OOO angles, where the larger systems permit greater configurational flexibility and, hence, produce a wider range of angle distributions. The OO radial and OOO angle distributions have a critical role in proton transfer pathways. Also see Appendix E.

To quantify the degree to which the fragment structures differ between the two systems, we first gauge the distribution of data around the training set centroids.

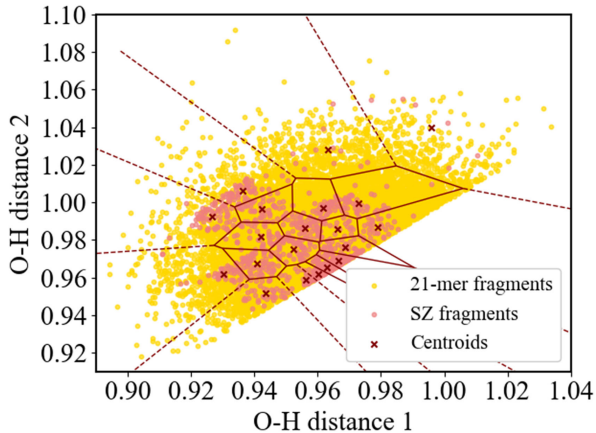


FIG. 15. Distribution of data for the water fragments and corresponding centroids obtained from k means. The centroids were derived from the solvated Zundel data (see Table I). Note that all data lie above the line where the two O-H distances are equal, since “O-H distance 1” is chosen to be the shorter distance. Clustering is conducted only on the two O-H distances for demonstration.

A simplified heuristic for such a data distribution is provided in Fig. 15 for water fragments found within both the solvated Zundel as well as the 21-water cluster. In Fig. 15, a significant difference in water fragment geometries can be observed by comparing the distribution of red and yellow dots. To further probe this difference, we consider the Cartesian distance between the descriptor vector for the k -means centroids and the data points from the two systems, that is, in some sense the level of confidence for a given data point given its proximity to a training point. We then compute the distribution of these Cartesian distances from each fragment to their closest training geometries (obtained from the k -means centroid geometries), and these are shown along the horizontal axes in Fig. 16. The distributions of distances for fragments from the two systems are displayed using two histograms in the top in Fig. 16. From the top of each figure in Fig. 16, a clear distinction between the distributions for the two systems begins to appear from H_4O_2 , which coincides with the pattern of MAE differences in Table II. As the size of the fragment increases, the difference in geometry distribution also increases as shown by the right-shifting peaks of blue histograms on the top in Fig. 16.

In the middle and bottom of each figure in Fig. 16, we present the error in target energy for 21-mer fragments and the error in CCSD energy for the solvated Zundel fragments, respectively, which results from the difference in the distribution of data noted above. From the scale of the vertical axes in the middle and bottom, the effect of this perturbation in distribution is already clear. This strongly indicates that, for geometries with large distances to the primitive training set, the predictive accuracy is generally low. Therefore, an amended transferring process needs to be designed to expand the valid training set regions to build confidence in the accuracy of predictions for all geometries.

C. The corresponding LLM challenge

The difficulty in extrapolating the protonated 21-water cluster surface, from the learning process set in place using solvated Zundel data, can also be understood from the LLM perspective presented in Sec. III. As noted in Fig. 9, the attention weights are limited by the breadth of context presented within the training data, as discovered during the tokenization and embedding processes.

To make the analogy much more precise, we posed the following question to the AI chatbot, Google Gemini [116]: “We wish for you to write a 186 word length paragraph, using only 51 words. However, to train you, we can only provide you lines that are precisely 51 words long.” Note that the number 51 and 186 in the above query arise from the fact that the solvated Zundel has 51 nuclear dimensions, whereas the protonated 21-water cluster has 186 nuclear dimensions.

Google Gemini provided the following response to this query. “This is a very specific and unusual constraint for

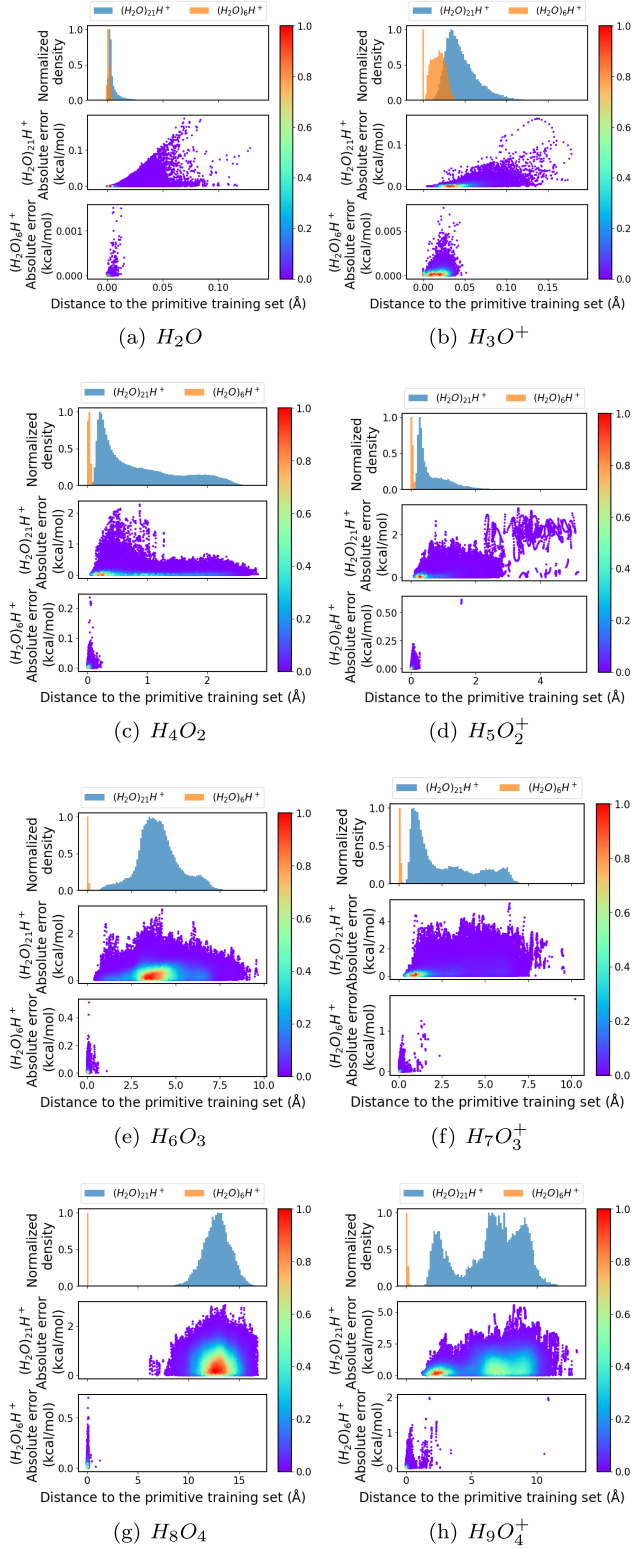


FIG. 16. Fragment data proximity from nearest training point (horizontal axis) and respective distributions (top). The absolute error is computed from the neural network predictions trained on solvated Zundel fragments. Color bars represent relative density.

training a language model, as typical training involves a wide variety of sentence lengths and structures. Here is what that restriction implies for how I would be ‘trained’ or what I could ‘learn’ from such data: (a) Limited grammatical diversity: If every sentence is precisely 51 words, I will not be exposed to the natural variations in sentence length and complexity common in human language. This could make it difficult for me to generate shorter, longer, or more nuanced sentences. (b) Focus on word choice within constraint: My ‘learning’ would heavily focus on how to use words effectively to reach that 51-word count, rather than focusing on natural flow or conciseness. (c) Potential for repetitive structures: To meet the word count consistently, the provided training sentences might employ similar grammatical structures or rely on a lot of descriptive phrases to extend length. I might then mimic these patterns in my own output. (d) Challenges in generalization: If I am only trained on 51-word sentences, my ability to understand and generate text outside of that specific length might be severely hampered. I might struggle with prompts that require very brief answers or very detailed, extended explanations. In essence, while I could still process and analyze the patterns within those 51-word sentences, my understanding and generation capabilities would be highly specialized to that very specific input format. It would be like learning to play a song with only one instrument and then being asked to perform a whole orchestra.”

Clearly, from an LLM constraint, this is a severe restraint; however, there is additional information that we possess, making it possible for us to take such a step in this paper, as we see in the next section.

V. A TRANSFORMER LEARNING PLATFORM TO REFINE THE FRAGMENT NEURAL NETWORKS TOWARD THE E^{target}

As discussed above, it is necessary to expand the previously defined training set regions. This is because the fragments sampled from the 21-mer trajectory include a much broader and more diverse range of geometries compared to those from the solvated Zundel trajectory. Also see the discussion in Appendix E in this regard. The difference in range of geometries sampled can be seen by comparing the geometry distribution of the water molecules represented by red and yellow dots in Fig. 15. Therefore, we expand the training sets using the portion of the 21-mer data that are not visited in solvated Zundel trajectory. The process produces an adapted family of neural networks that apply to both the solvated Zundel data and 21-mer geometries. For this purpose, we introduce a systematic process to incrementally expand the domain of training beginning from the data space of fragment geometries generated using the solvated Zundel AIMD trajectory. The process is illustrated in Fig. 17.

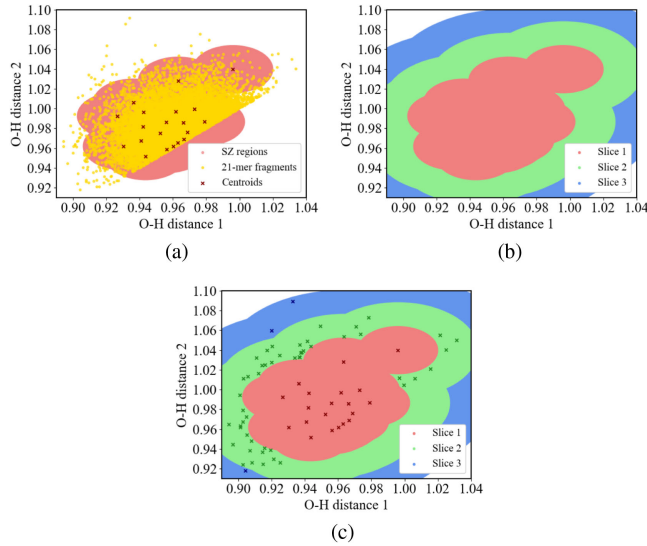


FIG. 17. (a) We place a circle with on each centroid (the same as in Fig. 15) with radius computed from the maximum sample-to-centroid distance for solvated Zundel fragments to define training regions. These regions imply that any data involved can be accurately predicted by solvated Zundel models. (b) To include data outside the training regions, we further slice the space based on their distance to the closest centroid and create an incremental transfer process. (c) Centroid positions for slices 2 and 3 are obtained from the recursive minibatch k means. (a) Training regions as the union of circles centered at centroids. (b) Slicing of the space around training regions for transferring. (c) Centroid positions.

We first imagine the full data space for each fragment to be divided into two orthogonal dimensions. One dimension is well sampled during the solvated Zundel trajectory, but configurations in the second dimension are sampled only during an AIMD trajectory for the larger water cluster system. In summary, our goal is to facilitate the sampling of this second orthogonal dimension, since the first dimension is already well sampled during the solvated Zundel study. See Figs. 10 and 12. This is achieved by first interpreting the entire orthogonal data space as one single effective dimension and tessellating this dimension into a series of slices based on the distance to the closest reference points from the solvated Zundel training set. For example, in Figs. 17(a) and 17(b), the red region represents the data span of the H_2O fragments obtained from the original solvated Zundel simulation (compare with Fig. 15). Expanding this space into the orthogonal 21-mer data space involves creating multiple additional layers or “shells” in that data space surrounding the solvated Zundel core displayed in red. The additional slices incrementally added are shown in green and blue in Fig. 17(b). These slices facilitate an incremental learning procedure to go beyond the neural networks obtained from the solvated Zundel systems based on the proximity of data to inner layers. In Fig. 17, the slice width (that is, the width of the

green and blue regions) is equal to the radius of the red circles, which represent the maximum sample-to-centroid distance for each training sample in the solvated Zundel data. (Further details on this geometric tessellation procedure of the data space are given in Appendix B.) Following this, for each new slice of the space (that is, for example, the green and blue slices in Fig. 17), we apply recursive minibatch- k -means algorithms [114] (a geometric tessellation procedure for higher-dimensional spaces) on geometries within the specific slice to identify a set of geometries with desired mutual distances to represent the slice of fragment energy space. This algorithm aims to perform k -means clustering recursively to reduce complexity and automatically derive an appropriate number of clusters and training samples to facilitate description of the protonated 21-mer surface. More details about the recursive minibatch- k -means clustering algorithm can be found in Appendix F. Figure 17(c) shows additional centroids in each slice. By comparing Figs. 17(c) and 15, we note that the additional centroids in the green and blue slices in Fig. 17(c) effectively sample the region in yellow in Fig. 15 and generate additional training samples to adapt the neural networks from the solvated Zundel subspace to those seen in larger clusters.

After identifying the additional centroids using the recursive k -means clustering algorithms discussed in detail in Appendixes B and F, an additional training set is formed from the geometries closest to each centroid to adapt the neural network models. Once the neural networks are trained on the additional training set, they are retrained with geometries closest to all existing centroids as a fine-tuning process. The complexity of neural network training is reduced here, because the transfer process is conducted based on the proximity of the data to the primitive training set and then adapted gradually to new regions. In the next section, we demonstrate the efficacy of this algorithm by gauging the error in the protonated 21-mer surfaces when these algorithms are implemented.

VI. PROTONATED 21-WATER CLUSTER POTENTIAL ENERGY PREDICTIONS AFTER INCREMENTAL TRANSFORMATIONS IN SEC. V

A. Fragment accuracy improves during the transfer learning protocol introduced in Sec. V

To improve our neural network models for protonated 21-mer, in Sect. V, we discussed how the full data space is divided into two dimensions. One of these dimensions contain fragment geometries in close proximity to training data obtained from the solvated Zundel system, and the other includes fragment geometries that are farther away from these original training data. We presented these as two orthogonal learning dimensions in Sec. V. We also discussed how the second dimension is partitioned, incrementally, into a series of slices that forms subspaces as shown in

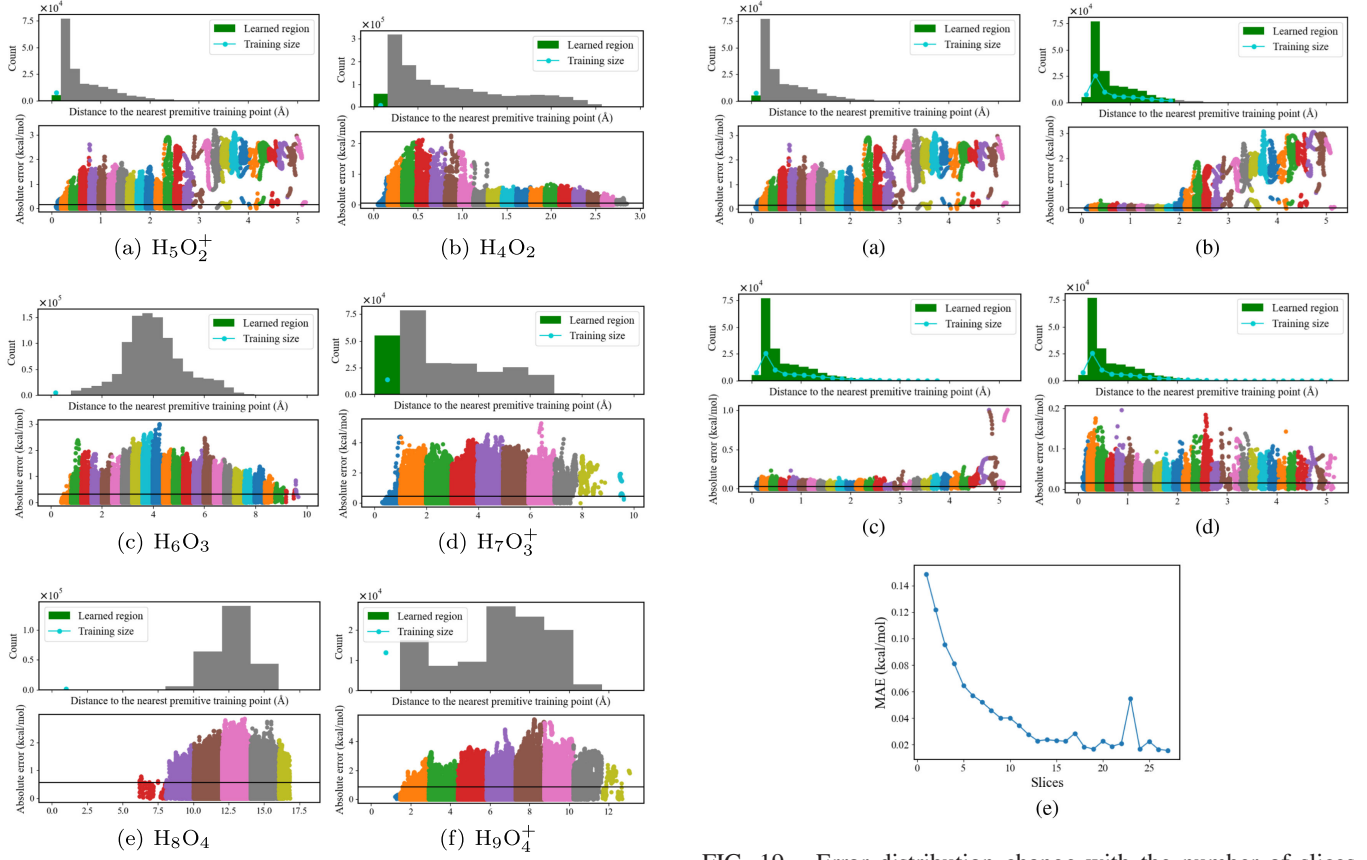


FIG. 18. Geometry distributions (top) and error distributions (bottom) for (a) H_5O_2^+ , (b) H_4O_2 , (c) H_6O_3 , (d) H_7O_3^+ , (e) H_8O_4 , and (f) H_9O_4^+ . Fragment structures deviate significantly between the solvated Zundel and 21-mer trajectories necessitating the transformer algorithm in Sec. V. The MAEs are noted using dark horizontal lines in the bottom, and these are also reported in Table III under the “MAE before transfer” column. Compare these MAEs with the solvated Zundel MAEs in Table II along with the increased probabilities of these fragments in 21-mer as seen from the $\omega_{f,r,\mathcal{R}}$ in Table II.

Fig. 17(b). We now apply this partitioning scheme for all fragments. Figures 18 and 19 complement our discussion here, with additional data for all other fragments in Appendix G. The said partitioning of the orthogonal dimension is presented in Fig. 18, top, where the horizontal axes refers to the distance to the nearest training set data point (which was obtained as above from the solvated Zundel data). It is this horizontal axis that is taken as the orthogonal dimension as referred to in Sec. V and sliced into smaller regions where additional minibatch- k -means clustering calculations are performed. The vertical axis in Fig. 18, top, shows the number of geometries inside the corresponding slice. In gray we show the distribution of geometries obtained from the protonated 21-water cluster trajectory, and in green we show the data space that was already visited in the solvated Zundel trajectory. Clearly, even for H_5O_2^+ , there is a large deviation of structures

FIG. 19. Error distribution change with the number of slices trained for H_5O_2^+ . (a) 1 slices learned. (b) 10 slices learned. (c) 20 slices learned. (d) All 27 slices learned. (e) MAE changes with number of slices transferred.

obtained from the 21-mer trajectory, as compared to those obtained from the solvated Zundel trajectory, and these new structures contribute significantly to the overall errors as seen from the bottom in Fig. 18. Comparison of Fig. 18, top, to Fig. 15 shows how dramatically the structures deviate as the fragment sizes grow. In Appendix E, we analyze these deviations in structures sampled between the solvated Zundel and 21-mer data. The goal of the incremental slicing algorithm presented in Sec. V is to improve on these errors, as we discuss below.

Details regarding the construction of the histogram error plots are as follows. The slice widths, or bin sizes in Figs. 18 and 19 and complementary figures in Appendix G, are determined from the maximum sample-to-centroid distances from the 10% minibatch- k -means clustering. This is identical to the sampling process to generate the primitive training set on solvated Zundel datasets as we discussed in Sec. IV. This distance approximates the farthest sample that can be accurately predicted by the model, thus serving as a boundary between the trained and untrained regions of the potential energy data space. Consequently, we plot the regions trained from solvated Zundel in green in Figs. 18 and 19. Each bin in the histogram plots in the top also includes a cyan dot to

TABLE III. The third column shows the number of fragments from the protonated 21-water cluster trajectory used for improving the accuracy of the original neural networks developed using the solvated Zundel trajectory dataset. The fragment mean absolute errors before and after transferring are also shown. Note the effect on the larger clusters.

	$\text{H}_{13}\text{O}_6^+$ training set size	Approximate transfer size ^a	$\text{H}_{43}\text{O}_{21}^+$ dataset size		MAE before transfer ^b	MAE after transfer ^a
H_2O	3891	16 000	559 877	3%	0.00	0.00
H_3O^+	1704	4000	34 297	11%	0.00	0.00
H_4O_2	6238	480 000	1 356 707	35%	0.06	0.00
H_5O_2^+	7751	76 000	190 727	40%	0.15	0.02
H_6O_3	4694	470 000	1 014 960	46%	0.31	0.02
H_7O_3^+	13 958	90 000	257 260	35%	0.45	0.07
H_8O_4	1575	134 000	253 889	53%	0.59	0.05
H_9O_4^+	12 414	65 000	108 123	60%	0.88	0.10

^aNumbers show the order of magnitude of data.

^bIn kcal/mol.

indicate the number of training samples used for that specific slice in the incremental training process. This helps to assess the extent of training needed as compared to the total number of samples in each slice (i.e., the respective bin heights). Notably, in the first bin, the number of training samples exceeds the number of fragments from the protonated 21-water cluster trajectory, which arises from the limited fragment geometric overlap between the solvated Zundel and the protonated 21-water cluster trajectory.

Hence, we incrementally create additional training samples from each slice of the space (as discussed in Sec. V) and adapt the neural networks. The resulting improvement on prediction error can be observed by comparing Figs. 19(a) and 19(b). In the bottom in Fig. 19(b), a clear distinction on error distribution emerges between trained and untrained slices. Predictions remain accurate across the ten learned slices, while errors in the untrained slices remain high. As the incremental transfer process continues, the error in systematically reduced, from a maximum value of approximately 3 kcal/mol in Fig. 19(b), to a maximum value of 1 kcal/mol in Fig. 19(c), and finally to 0.2 kcal/mol in Fig. 19(d). This is also the case for larger fragments as seen from the figures in Appendix G. The systematic reduction in mean absolute errors due to this incremental learning process is summarized in Fig. 19(e) and the corresponding final figures in Appendix G. These final errors are also summarized in Table III and are comparable to the errors from primitive neural networks trained only on solvated Zundel fragments and reported in Table II. At this point, the adapted neural network models have successfully evolved to make accurate predictions for both solvated Zundel and protonated 21-water cluster systems. In the next section, we integrate all adapted models to perform full system predictions and analysis for the protonated 21-water clusters.

B. Full system potential energy prediction accuracy after transferring all fragment models

Upon incremental modification of all fragment neural networks, the overall error in the protonated 21-water

cluster reduces from 5.32 kcal/mol [the error when the solvated Zundel models are directly used, that is, $\{\Delta E_{\alpha,r}^{\text{ML},1}(\bar{\mathbf{x}}_{\alpha,r})\} \rightarrow E_{\text{system},2}^{\text{target}}(\bar{\mathbf{x}})$, Fig. 12] to 0.36 kcal/mol [when the modified, incremental, and directed learning process is used, that is, $\{\Delta E_{\alpha,r}^{\text{ML},2}(\bar{\mathbf{x}}_{\alpha,r})\} \rightarrow E_{\text{system},2}^{\text{target}}(\bar{\mathbf{x}})$]. The corresponding distribution of errors is seen in Fig. 20, and a summary for $\mathcal{R} = 3$ before and after transferring is presented in Fig. 21. Importantly, the distribution of errors shrinks toward the main diagonal with increase in $\mathcal{R}$ in Fig. 20, unlike the case in Fig. 12, and the predictions remain accurate for the entire range of structures. This can also be seen in Fig. 21. In Fig. 21(a), the error distribution before transferring exhibits a peak near 3 kcal/mol, corresponding to the purple region representing the highest density in Fig. 12(d). After neural network modification, the peak shifts reflect the improved accuracy displayed in Fig. 20(d), where the purple region is strictly along the diagonal.

We summarize all mean absolute errors in Fig. 22. These are computed with respect to the graph energy at $\mathcal{R} = 3$, and, hence, the blue bar on the right in Fig. 22 is identically zero for $\mathcal{R} = 3$. The orange ML error very clearly shows the deviations predicted in Fig. 12 that are then corrected after modifying the fragment neural networks as seen from the green histograms. We also compare the corresponding computational cost for constructing these models with direct computation of CCSD energies in Sec. VIC.

As per Table III, one may expect major improvements in the protonated 21-water cluster potential energy predictions by including the larger $\mathcal{R} = 2$ and $\mathcal{R} = 3$ fragments, H_6O_3 , H_7O_3^+ , H_8O_4 , and H_9O_4^+ . Therefore, we examine the potential energy predictions for the protonated 21-water cluster when the transformed neural networks are used only for larger fragments. The error is displayed in Fig. 23, where the first blue bar corresponds to the MAE from the green bar for $\mathcal{R} = 3$ in Fig. 22. As seen from comparing the blue and orange bars in Fig. 23, omitting H_2O and H_3O^+ fragments from the modification process leads to negligible differences in MAE. This result is also consistent with Table II. However, excluding H_4O_2 results in an increase in

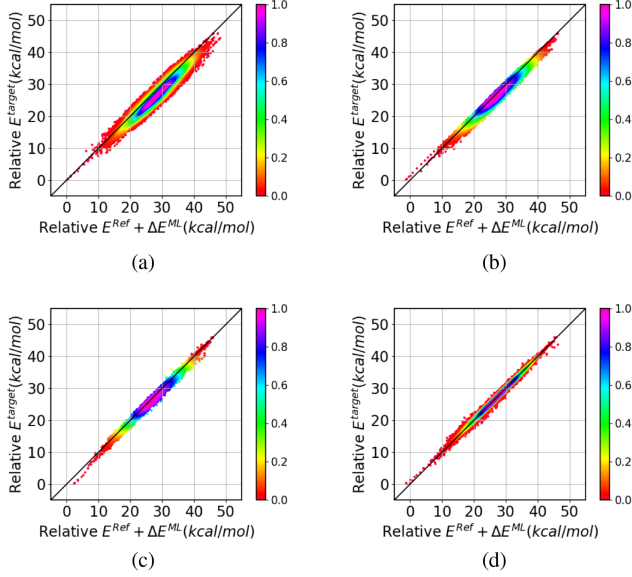


FIG. 20. The protonated 21-water cluster potential energy prediction accuracy from transformed NN models. (a) Using transformed NN models, $\mathcal{R} = 0$. (b) Using transformed NN models, $\mathcal{R} = 1$. (c) Using transformed NN models, $\mathcal{R} = 2$. (d) Using transformed NN models, $\mathcal{R} = 3$.

MAE from 0.36 to 0.65 kcal/mol as shown by the green bar in Fig. 23. This is because, although H_4O_2 has a low error of 0.06 kcal/mol when the solvated Zundel fragment neural networks are used, the population $\omega_{f,r}$ values in Table II are large. Thus, excluding these fragments in Fig. 23 has a significant effect. A similar behavior results when H_5O_2^+ is excluded, as shown by the red bar in Fig. 23 due to the larger error from these fragments as seen in Table II that is significantly reduced due to the modification of neural networks for these fragments as seen from Table III.

To further investigate the contribution of H_4O_2 and H_5O_2^+ to the total error in Fig. 24, we provide the distribution of energies for the protonated 21-water cluster when these two species are excluded from the transferring process. The corresponding mean absolute errors are represented by the green and red bars in Fig. 23. A

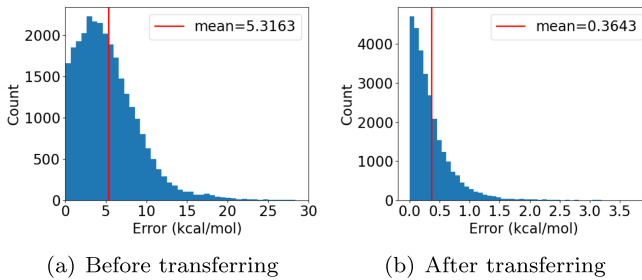


FIG. 21. The protonated 21-water cluster full system potential energy absolute error predicted from models before and after transferring. (a) Recasts the same data in Fig. 12(d), and (b) recasts the same data in Fig. 20(d) for easy comparison.

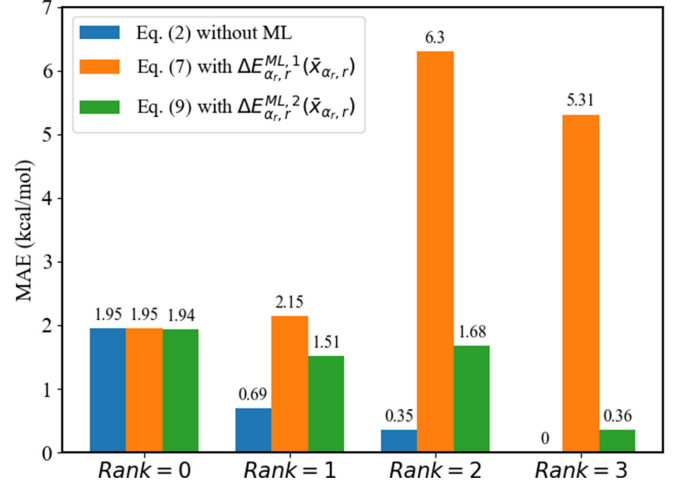


FIG. 22. Improved performance of the transformed neural networks for protonated 21-water cluster potential energy predictions. The blue histogram represents Eq. (2) without the use of machine learning. The orange histogram represents Eq. (9) but with $\{\Delta E_{a,r}^{ML,1}(\bar{x}_{a,r},r)\}$, that is, the solvated Zundel network models used to obtain $E_{\text{system},2}^{\text{target}}(\bar{x})$, and the green histogram depicts Eq. (9) as written, that is, $\{\Delta E_{a,r}^{ML,2}(\bar{x}_{a,r},r)\} \rightarrow E_{\text{system},2}^{\text{target}}(\bar{x})$. All MAEs are computed with reference to graph energy at rank 3.

comparison between Figs. 24 and 20(d) reveals that the error distribution has grown broader and more diffuse in Fig. 24. At the same time, the broadened error distribution displays a symmetric nature along the central diagonal. This observation suggests that the errors associated with H_4O_2 and H_5O_2^+ may have a probabilistic nature and, thus, perhaps less complicated as compared to off-axis shifts and deformations in Fig. 12(d). Thus, future work may benefit from stochastic methods to further reduce the amount of training samples required for such fragments.

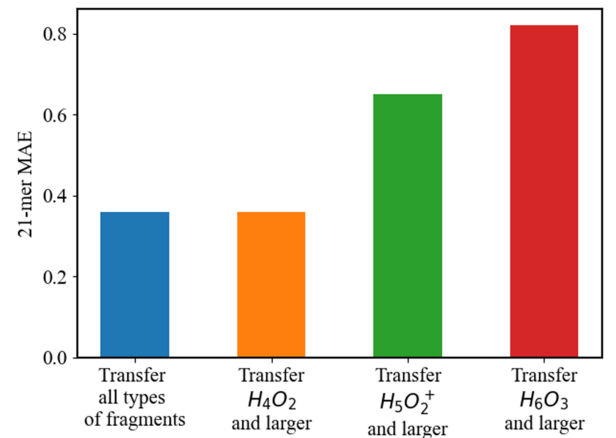


FIG. 23. The protonated 21-water cluster MAE if the transformation process is conducted only on fragments larger in size than those listed. This shows that nodes do not need to be transformed and accuracy may be maintained even when all edges are not transformed.

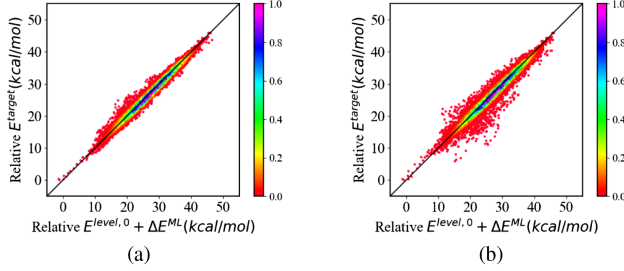


FIG. 24. The protonated 21-water cluster potential energy error distribution similar to Fig. 20, but only larger fragments are transferred as discussed in Fig. 23. (a) Transfer H_5O_2^+ and larger. (b) Transfer H_6O_3 and larger.

C. Computational cost associated with graph-theoretic neural network models for post-Hartree-Fock accuracy

Here, we summarize the total cost for constructing neural network potential energy surfaces and compare it with direct evaluation of fragment CCSD energies and full system CCSD energies. The total training cost reported here includes both fragments obtained from the solvated Zundel trajectory as well as those appended through incremental orthogonal training as dictated by the geometries found in the 21-mer trajectory. The resultant training data represent approximately 32% of all fragment geometries obtained from the solvated Zundel and protonated 21-water cluster trajectories.

Since the goal is to reduce the computational effort to allow higher-quality electronic structure, we discuss the formal computational complexity. In order to quantify the discussion, we assume that the target molecular system is divided into $\mathcal{N}$ identical nodes each with basis set size of $\mathcal{F}_0$ which determines the computational complexity for the molecular fragment associated with each node. Thus, the size of basis set for the full system is roughly $[\mathcal{N} * \mathcal{F}_0]$. A computationally sequential (or serial) treatment of the problem leads to evaluation of each rank- r term at two levels of theory. If we further assume that the Ref calculation scales as $\mathcal{O}(N^{L_0})$ and target scales as $\mathcal{O}(N^{L_1})$, then the overall target calculation scales as $\mathcal{O}((\mathcal{N}\mathcal{F}_0)^{L_1})$. The fragment calculations then scale as

$$\begin{aligned} \text{Scaling}_{\text{Eq. (4)}} &\approx \sum_{r=0}^{\mathcal{R}} \sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r, r}^{\mathcal{R}} \mathcal{O}((r\mathcal{F}_0)^{L_1}) \\ &= \sum_{r=0}^{\mathcal{R}} \mathcal{O}((r\mathcal{F}_0)^{L_1}) \left[\sum_{\alpha_r \in \mathbf{V}_r} \mathcal{M}_{\alpha_r, r}^{\mathcal{R}} \right] \\ &\approx \mathcal{O}((\mathcal{R}\mathcal{F}_0)^{L_1}) N_{\mathcal{R}}, \end{aligned} \quad (29)$$

where $N_{\mathcal{R}}$ is the number of rank- $\mathcal{R}$ fragments. The calculations here are performed in an asynchronously parallel fashion [48], and, hence, the scaling in Eq. (29) is further reduced. The serial version of the expression in Eq. (29) is shown using the orange histogram in Fig. 25, whereas the

quantity $\mathcal{O}((NM)^{L_1})$ is represented in red for the protonated 21-water cluster system. As can be seen, there is significant reduction in computational effort even in the absence of exploiting parallelism. The histogram in blue represents the portion of the orange histogram that is used for training here in constructing the machine-learning models. As noted above, this is roughly 32% of the orange histogram and also does not represent the available parallelism in developing neural networks simultaneously for multiple fragments.

While the discussion above provides general guidelines and argues the power of our approach, we next provide more precise scaling costs that accurately present the systems treated here. To estimate and compare the costs, we use CCSD scaling cost [117] of N^6 for N electrons, and this represents the item shown in red in Fig. 25 for all 21-mer data. Based on this scaling, the protonated 21-water cluster has 211 electrons, leading to a total computational cost of $211^6 * 28\,294 = 2.5 * 10^{18}$ for all 28 294 geometries in the trajectory. This value is represented by the red bar in Fig. 25 and normalized as unity for ease of comparison.

To evaluate the cost for all fragment energies obtained from the protonated 21-water cluster trajectories, we take the number of fragments of each kind listed in Table I, as already seen from Eq. (29), and multiply by the factor $\mathcal{F}_r^6$ for $\mathcal{F}_r$ electrons for rank- r fragments. For example, we found 559 877 water fragments each with ten electrons in the protonated 21-water cluster trajectory. Then, the cost for all water fragments becomes $10^6 * 559\,877 = 5.6 * 10^{11}$. Similarly, there are 34 297 H_3O^+ geometries found, each with 11 electrons. The total cost for all H_3O^+ becomes $11^6 * 34\,297 = 6.1 * 10^{10}$. When we repeat the same process for all types of fragments listed in Table I, the total

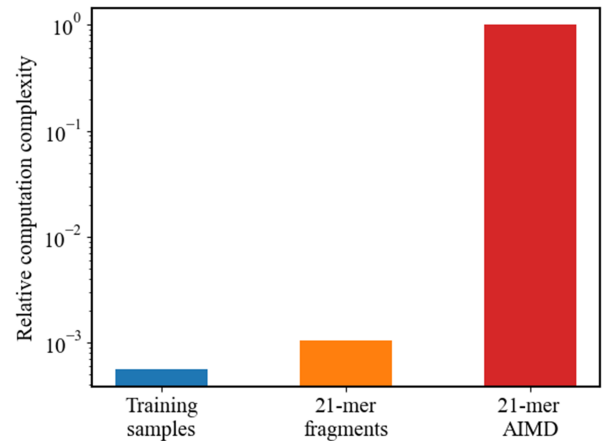


FIG. 25. Relative computational cost for CCSD energies for all training samples, fragments from the protonated 21-water cluster trajectory, and the full system potential energies of protonated 21-water clusters. Training samples shown in the blue bar involve both primitive and additional training sets. Fragments shown by the orange bar are produced from the protonated 21-water cluster trajectories.

fragment cost is the sum of cost for each type of fragments as $5.6 * 10^{11} + 6.1 * 10^{10} + \dots = 2.6 * 10^{15}$. The total fragment energy cost is 3 orders of magnitude smaller than the full system cost, indicating a significant computational saving achieved by applying Eq. (2). This is because the largest major fragments obtained from the protonated 21-water cluster are H_9O_4^+ , which contains only 41 electrons, far fewer than the 211 electrons in the full systems.

Similarly, the blue bar in Fig. 25 illustrates the total cost of all fragments used for training our neural network models, which involve both primitive and additional training sets obtained from the two trajectories. For example, for water fragments, as can be seen from Table III, we use 3891 geometries from the solvated Zundel trajectory for training an initial set of neural network models as described in Sec. IV. Another 16 000 water geometries from the protonated 21-water cluster trajectory are included as the additional training set to adapt the neural networks to the orthogonal regions of the fragment space. Then, the total training set cost for the water neural network model is computed as $10^6 * (3891 + 16000) = 2.0 * 10^{10}$. We again sum over all fragments that are modeled by independent neural networks and plot as the blue bar in Fig. 25. These results provide an intuitive demonstration of how Eq. (17) effectively reduces the prohibitive computational cost for direct full system mapping $\mathcal{Q}$ to a family of much smaller and tractable mappings $\{\mathcal{Q}_{\alpha,r}\}$.

VII. CONCLUSION

The cost of acquiring sufficient training samples to construct accurate neural network models for large system potential energy surfaces at the coupled-cluster level of theories has presented a great challenge to computational chemistry. In this paper, we present a transfer learning protocol which extends from a smaller known (learned) system, in our case the solvated Zundel system, to predict the potential energy surface for a much larger system, in our case the protonated 21-water cluster. Thus, we construct learning models appropriate for a 51-dimensional space (solvated Zundel) and adapt those models to extrapolate the quantum mechanical energies for a 186-dimensional space (the protonated 21-water cluster).

Toward this, we begin by using our previously developed graph-theoretic molecular fragmentation procedure for configuring common fragments from both systems. Then, we construct a set of neural network models that learn on the energies from each type of fragment separately from a solvated Zundel library of structures. To better represent the graph-theoretically generated fragment geometries and understand their differences, we designed a permuted interatomic distance vector that follows translational, rotational, and permutation invariance as the descriptor. These neural network models provide highly accurate predictions on the solvated Zundel potential energies at an MAE of 0.36 kcal/mol but lose accuracy

when directly used for predicting energies for the protonated 21-water cluster with an MAE of 5.32 kcal/mol due to the change in distribution of fragments structures. Thus, we design a modified, incremental, and directed transfer learning process that first distinguishes the extent of fragment structural differences between the two systems. Once these differences are understood, an incremental learning process yields results with sub-kcal/mol accuracy of the full potential energy surface of the protonated 21-water cluster. We obtain highly accurate potential energy predictions for a protonated 21-water cluster dynamics trajectory, with MAE of 0.36 kcal/mol at CCSD level while also maintaining good accuracy for the initial solvated Zundel system using the same family of neural network models.

The graph-based representation provides a robust foundation for treating different fragments independently, making a key deviation from the architecture used in transformer-based large language models. In transformers, all sentence components are processed uniformly through multiple sequential layers of attention computation, resulting in a deep stacking architecture. In contrast, our method creates a distributed set of independent neural network blocks, corresponding to individual fragment species. We consider this architectural distinction an advantage that allows us to focus on the more critical components of the system. Here, importance not only refers to the attention score of a word or the multiplicity of a molecular fragment, as discussed here, but also includes the chemical significance of a fragment by incorporating many-body interactions.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation Grant No. CHE-2102610 (Creativity Extension) to S. S. I. The computational facilities at Indiana University are duly acknowledged and have been critical to this effort. Acknowledgment is due in part to the Lilly Endowment, Inc., for their support of the BigRed computing facility at Indiana University widely used in the effort represented in this publication. This work was also supported in part by Shared University Research grants from IBM, Inc., to Indiana University, which supports the Scholarly Data Archives.

APPENDIX A: ILLUSTRATION OF FIGS. 4 AND 5 FOR BONDED AND NONBONDED SYSTEMS

In Refs. [41,42,46], we have used the partitioning scheme in Sec. II to study both bonded as well as nonbonded systems. In these studies, coupled-cluster accuracy is obtained at DFT cost from the direct use of the graph-theory approach discussed in Sec. II. We illustrate these here for completeness. In Ref. [41], we divide a protein fragment, a poly-alanine chain shown in Fig. 26, into nodes, edges, faces, and higher-rank objects, where each amino acid fragment is chosen as a separate node fragment. As a result, edges are dimers, faces are trimers

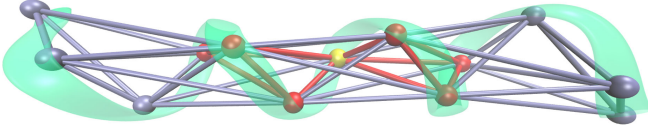
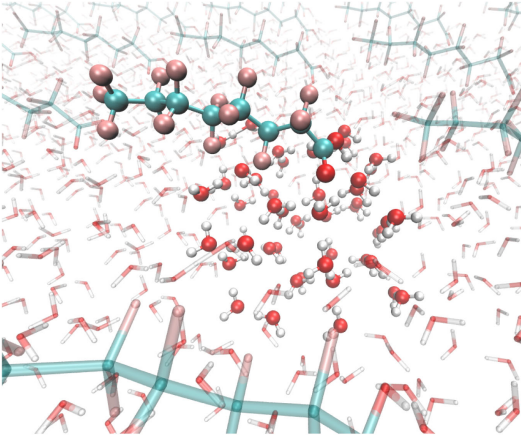
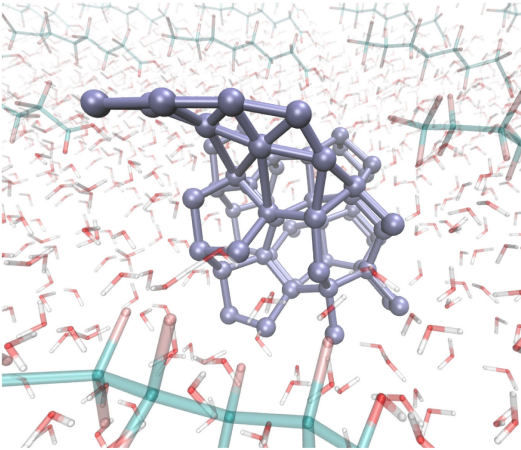


FIG. 26. In Refs. [41,42], a poly-alanine system is divided into simplicial complexes as discussed in Sec. II. This is a bonded system, and here nodes are alanine fragments. When bonds are cut to create fragments, these terminated using link atoms to satisfy valency as done within the ONIOM protocol of mechanical embedding [85]. The figure shows higher-order simplexes and, for example, the node in yellow as interactions that are part of several fragments.

(bonded and nonbonded), and so on. Likewise, in Ref. [46], we consider a solvation study of perfluoro-octanoic acid, an environmental pollutant, on a periodic slab of water. Here, the partitioning is complicated by additional interactions, and



(a)



(b)

FIG. 27. Similar to Fig. 26 but for a heterogeneous, solvated system [46]. Interactions between the organic substrate and water are captured at multiple levels through the use of simplicial complexes. (b) Shows the graph representation of (a), where higher-order simplexes are clearly seen.

larger-rank simplexes involve interactions of the organic system with water. The scheme is presented in Fig. 27. In all cases, when chemical bonds are broken to create fragments, the dangling bonds are passivated through link atoms as is routinely done in the ONIOM protocol [85]. This kind of passivation is called mechanical embedding.

APPENDIX B: MINIBATCH- k -MEANS SAMPLING AND INITIAL NEURAL NETWORK MODEL SETUP FOR THE SOLVATED ZUNDEL

To generate the initial neural network model for solvated Zundel fragments, we begin by collecting all fragments from the solvated Zundel trajectory and create a data bank. For each type of fragment, we use minibatch k means for 10% sampling. The minibatch- k -means algorithm is a variant of the clustering algorithm k -means clustering, which essentially tessellates or divides the data space into k mutually exclusive regions called clusters $\{C_j\}$. Every cluster is represented by a centroid $\bar{\mathbf{r}}_j$ which is computed as the arithmetic mean of all data points $\{\mathbf{r}_i\}$ when these are assigned to the closest centroid. The quantities $\{\mathbf{r}_i\}$ refer to interatomic distances within fragments, and these are specifically ordered as per the recipe described in Appendix C. The k -means algorithm aims to find a preset number of centroids or clusters iteratively to minimize the cost function:

$$\sum_{j=1}^k \sum_{\mathbf{r}_i \in C_j} |\mathbf{r}_i - \bar{\mathbf{r}}_j|^2. \quad (\text{B1})$$

The minibatch- k -means algorithm provides an efficient approach by using only a random subset of data (known as a batch) to update the centroid positions during each iteration. After finding the centroid positions, we construct the primitive training set as the set of closest data points to every centroid:

$$\{\mathbf{s}_j\} = \{\mathbf{r}_i | \min_{\mathbf{r}_i} |\mathbf{r}_i - \bar{\mathbf{r}}_j|\}. \quad (\text{B2})$$

The data points $\{\mathbf{s}_j\}$ are now the closest data points to each centroid $\mathbf{r}_i$ from the closest centroid $\bar{\mathbf{r}}_j$ and represents our training data points represented below as $x_{n=0}$.

In this paper, we place two neural networks in each model as the array, and each neural network is configured with four hidden layers. Each hidden layer consists of the number of nodes as 4 times the number of features or the number of interatomic distances introduced in Appendix C. The activation function on hidden layers has been modified to a Gaussian function:

$$x_{n;a} = \exp[-(w_{n-1,n;a} \cdot x_{n-1})^2], \quad (\text{B3})$$

where $x_{n;a}$ is the a th element of the layer vector x_n and $w_{n-1,n;a}$ represents the a th row of the weight matrix $w_{n-1,n}$

connecting layers n and $n - 1$. The quantity $(w_{n-1,n;a} \cdot x_{n-1})$ is the dot product of the a th row of the weight matrix $w_{n-1,n}$ with the vector x_{n-1} corresponding to the $(n - 1)$ th layer. Thus, each hidden layer vector is an exponent of an elementwise square of the previous layer vector. During the training process, a bias of “1” is appended to each vector in $\{\mathbf{s}_j\}$ from Eq. (B2) as the inputs for neural networks, and, hence,

$$x_{n=0} = \begin{bmatrix} \mathbf{s}_j \\ 1 \end{bmatrix}. \quad (\text{B4})$$

The bias with its associated neural network weights is used for optimizing the positions of Gaussian functions during training. The Gaussian-type function is useful for reducing the long-range effect during transferring so that the currently training slice does not interfere much with the previously learned slices of data space.

The neural network model that we used to fit the fragment energy pattern is a neural network array. This array consists of a series of neural networks, each of which learns on the cumulative error on predictions from prior neural networks in the series. So, for M neural networks in the series, the final prediction is

$$\Delta E^{\text{ML}} = \sum_m^M \Delta E^{\text{ML},m}. \quad (\text{B5})$$

APPENDIX C: A CONSISTENT, PERMUTATIONALLY INVARIANT, FRAGMENT MOMENT OF INERTIA FRAME OF REFERENCE FOR DESCRIBING FRAGMENT GEOMETRIES AND DESCRIPTORS

We intend to introduce a consistent framework for description of fragment geometries that arise from different molecular systems. Such a descriptor should allow for translational, rotational, and permutational invariance within the fragments, so that chemically identical fragments derived from different systems, or such fragments derived from different physical regions of the same system, have the same consistent mathematical description and, thus, can be compared with each other.

We begin with the interatomic distance matrix of fragments as descriptors but then recognize that the ordering of atoms between fragments may not be chemically consistent. For example, let us consider a water-wire fragment $[(\text{H}_2\text{O})_n]$ fragment which is found in multiple chemical systems and has a critical role in proton transfer in water clusters and in condensed phase systems. Hence, a fragment of this kind can be found in most water cluster systems. An example of such a system can be found in Fig. 28. This specific system contains eight protons, but all eight protons do not have the same chemical environment and properties. While five protons are on the periphery,

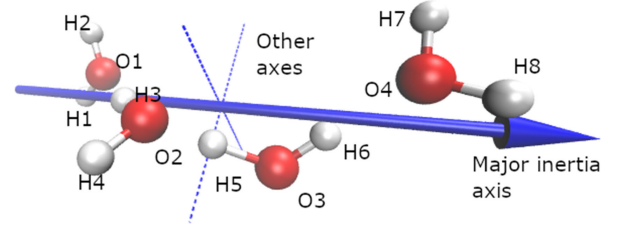


FIG. 28. An illustration of the moment of inertia frame of reference and the corresponding atom ordering for a four water-wire fragment. The major axis is the mass moment of inertia axis with the smallest absolute mass moment of inertia. The dashed lines represent the two other inertia axes.

three protons are shared between two oxygen atoms. Additionally, the shared protons themselves have different secondary environments. For larger water clusters, these differences become further complicated where our chemical description of the cluster must maintain a distinction between the shared protons on the outer layers of the cluster and those on the inner layers. To arrive at a consistent notation that retains the chemical meaning of each atom, we use the following steps.

- (1) We first rotate the fragment geometry coordinates into center of mass coordinates and mass moment of inertia frame of reference. In Fig. 28, the large blue axis refers to the axis of the major moment of inertia, that is, the smallest inertia value.
- (2) Following this, the atoms are ordered in increasing order of atomic weights and then
 - (a) increasing order of coordinate projection along the major moment of inertia axis.
 - (b) For cases where the major axis projected value is identical, the second major axis projection value is used to resolve the order of atoms.
- (3) We then arrive at an ordered list of atoms for any given fragment that consistently represents the chemical environment that the specific atom resides within.
- (4) After permuting all atoms based on the relative positions along the fragment moment of inertia axis, the distance matrix is computed and reshaped into a vector to define the fragment descriptor. These fragment descriptors are referred to as “ r_j ” in Appendix B.

This procedure establishes a permutation-invariant ordering of atoms within each fragment. This modification improves the consistency of fragment representations across different systems by better preserving geometrical symmetry. Consequently, the number of training samples required and the training cost are both reduced. The effect can be seen by comparing the geometry span defined by an average inertia given by

$$D = \frac{1}{N} \sum_i |\mathbf{r}_i - \bar{\mathbf{r}}|^2, \quad (\text{C1})$$

TABLE IV. The inertia span in Eq. (C1) with and without a permutation invariant reordering introduced by Appendix C. The D value is computed from Eq. (C1) for fragments obtained from the protonated 21-water cluster trajectory.

Fragments	D (Å) without reordering	D (Å) with reordering
H ₂ O	0.06	0.05
H ₃ O ⁺	0.21	0.19
H ₄ O ₂	2.88	2.06
H ₅ O ₂ ⁺	4.93	2.86
H ₆ O ₃	5.12	3.72
H ₇ O ₃ ⁺	6.97	4.14
H ₈ O ₄	7.00	5.55
H ₉ O ₄ ⁺	8.26	7.07

where $\mathbf{r}_i$ is the i th interatomic distance vector and $\bar{\mathbf{r}}$ is the average of all $\mathbf{f}_i$ for a total of N geometries. In Table IV, we list the D values for all fragments generated from the protonated 21-water cluster trajectory with and without this ordering. As we can see from the table, such an ordering effectively reduces the spanning for most larger fragments.

It must be noted that permutational invariance is part of a broader class of requirements that arise from group theory. Specifically, for systems that obey certain group structures, the neural network must also transform according to the irreducible representations of that group structure. This is essentially the goal of equivariant NNs [118,119]. In our work here, we use a pragmatic approach to enforcing permutational invariance. Equivariance is preserved in our case through the mass moment of inertia axis projection of coordinates for each fragment (obtained from the simplicial complex).

APPENDIX D: DISTINCTIONS BETWEEN CHALLENGES IN QUANTUM-CHEMISTRY-BASED POTENTIAL ENERGY SURFACES AND LLM MODELS

Although languages and molecules can be represented as graphs and may have a similar overarching strategy, the inherent differences between languages and molecules introduce challenges during different stages of the solution process. For molecules, constructing the graph is straightforward, and the weights can be calculated with well-defined inclusion and exclusion relationships. In contrast, for language, the relationship between words is more complicated and abstract. Their relationship depends not only on their relative position in sentences, but also on their meaning, coherence, and logical reasoning connections. Therefore, the attention is computed and optimized to understand the whole sentence through multiple layers of transformer architectures.

On the other hand, the primary challenge for molecules lies in obtaining accurate fragment-level information, specifically the fragment energies. Unlike words in language processing, which benefit from clear contextual information available from large and easily available

language datasets for training, fragment energy modeling requires extensive and computationally expensive post-Hartree-Fock calculations to generate training data. Thus, the neural network models used for fragment energy embedding often have limited accuracy and generalizability. For example, as we demonstrate in this paper, models trained on fragments from solvated Zundel cannot be directly applied to the same sets of fragments from the protonated 21-water clusters. This limitation motivates the development of our systematic approach to incrementally expand the training space for fragment energy modeling.

From another perspective, a key feature of the transformer architecture is its ability to iteratively refine the representation or embedding of each word or token based on its contextual environment (sentences). In each layer, the attention computed from Eq. (21) followed by neural network transformations and residual connections are used to update the query, key, and value vectors for the next iteration of attention computation. Therefore, the same set of words can play different roles in different sentences. This refinement process is philosophically aligned with our approach, following Eq. (8) to update fragment energy representations from $\Delta E_{a,r}^{\text{ML},1}(\bar{\mathbf{x}}_{a,r})$ to $\Delta E_{a,r}^{\text{ML},2}(\bar{\mathbf{x}}_{a,r})$ based on the full system configuration, allowing local fragment properties to evolve in response to the global molecular information.

APPENDIX E: PROTONATED 21-WATER CLUSTERS OPEN NEW HYDROGEN TRANSFERRING PATHS COMPARED TO SOLVATED ZUNDEL

Associated with the differences in distribution noted in Fig. 14, the fragment energy surfaces obtained from within the protonated 21-water cluster and solvated Zundel system also show critical differences. The H₅O₂⁺ fragments from the solvated Zundel database span the orange trace in the bottom in Fig. 29(a). As discussed in Sec. IV, the initial family of neural network models was constructed for this distribution of structures and faithfully represents the solvated Zundel system. However, when the protonated 21-water cluster system is considered, the H₅O₂⁺ fragments display the behavior in Fig. 29(a). The top in Fig. 29(a) shows the distribution of OO radial distribution function for the H₅O₂⁺ fragment geometries obtained from the protonated 21-mer system whose energies can be determined from the solvated Zundel database (orange) and those that cannot with suitable accuracy (blue). Thus, the blue distribution is the orthogonal counterpart which we incrementally learn from based on the discussion in Sec. V. While the difference may be thought to be marginal for the H₅O₂⁺, as seen from Table III, this still results in a 0.15 kcal/mol error due to the number of H₅O₂⁺ in the larger water cluster. This error is reduced to 0.02 kcal/mol, when the additional distribution (blue) is incrementally added to the neural networks created from the solvated Zundel database. These differences become even greater for the larger

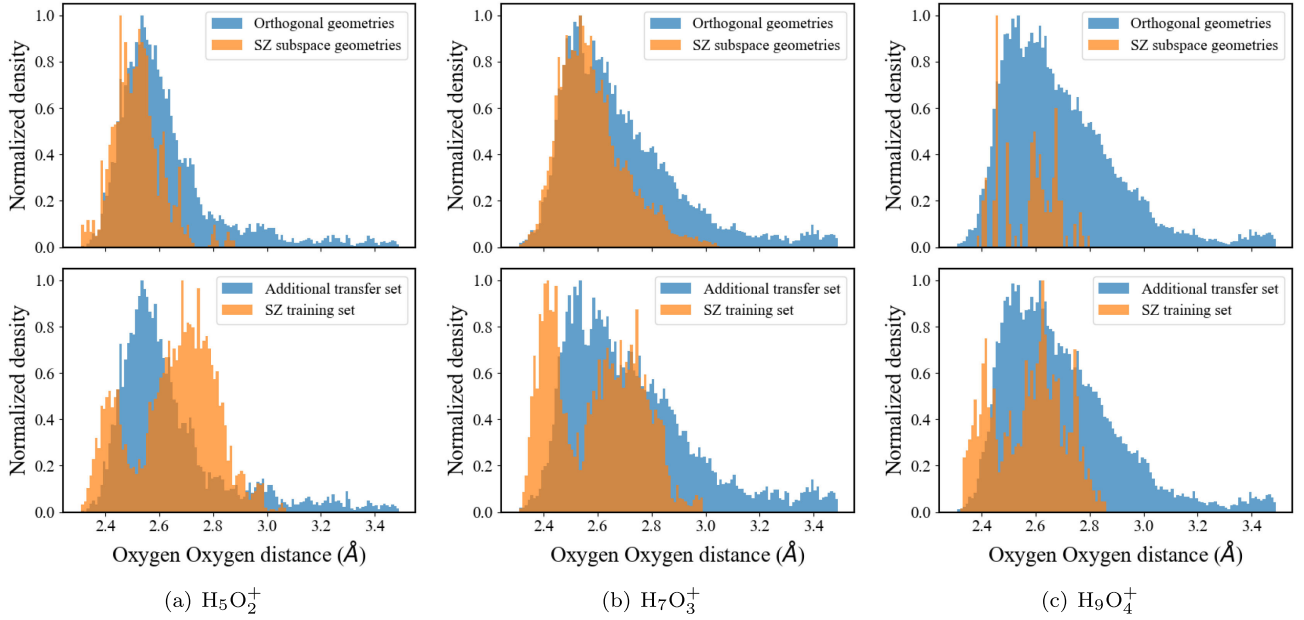


FIG. 29. The top shows the distribution of OO distances for the H_5O_2^+ (a), H_7O_3^+ (b), and the H_9O_4^+ (c) fragments from within the protonated 21-mer system. Specifically, the orange trace in the top includes those structures from the protonated 21-mer system that are accurately represented using the neural networks designed from solvated Zundel database. The blue traces in the top need additional incremental training using the methods discussed in the paper. Clearly, as the fragment size increases, the structures deviate from the solvated Zundel database in a big way, reflecting the behavior shown in Table III. For the bottom, the orange traces represent the pretrained structures from solvated Zundel, and the blue traces represent those that need additional training. As the cluster sizes grow, the deviation in OO distance sampled is clear. Since the OO distance is one major indicator of proton hops, the protonated 21-mer clearly samples a wider range of possibilities.

clusters, as seen for H_7O_3^+ [Fig. 29(b)] and H_9O_4^+ [Fig. 29(c)]. Thus, the larger clusters provide additional flexibility to the protonated fragments resulting in (a) more dynamical pathways for proton hopping and (b) distinct contributions to the overall energy of the system as seen from Fig. 23.

In Fig. 30, we present the OOO angle distribution which also clearly outlines the differences in geometries of the smaller fragments as the overall system size grows. The

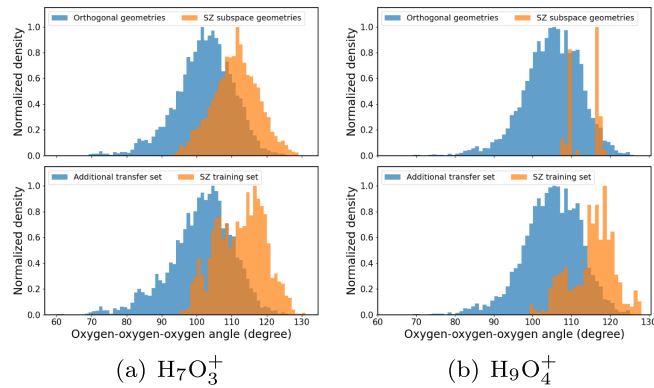


FIG. 30. Similar to Fig. 29, but OOO angle distribution for all connected trimer and quartermer subsystems. It is believed [94–97,120] that the surrounding hydrogen-bonded water molecules (around the protonated species in water) facilitate proton transfer, and in this figure we probe how this distribution changes with system size.

orthogonal incremental tessellation algorithm samples these complementary regions of data space. Figure 30(a) shows the distribution of OOO angles for the H_7O_3^+ , and Fig. 30(b) shows the same for H_9O_4^+ fragments from within the protonated 21-mer system. Clearly, these angle distributions deviate much more than the OO distributions in Fig. 29, reflecting the behavior already shown in Table III and the need for incremental training as discussed. Importantly, the larger spread of angles seen in Fig. 30 for the protonated 21-water cluster occur from secondary solvation shell interactions. These secondary solvation shell effects are present within our graph-based formalism.

APPENDIX F: THE RECURSIVE MINIBATCH- k -MEANS CLUSTERING WITHIN THE SLICES

The recursive minibatch- k -means algorithm intends to create clusters with sizes approximately the same as those from the 10% minibatch- k -means on the primitive data (slice 1). From Ref. [49] and Table II, we conclude that neural network models trained on the 10% training data, where the training data are obtained from k -means clustering, are highly accurate and represent solvated Zundel potential within sub-kcal/mol accuracy. Based on these training data for the primitive, solvated Zundel dataset, we compute the average inertia (η_0) to measure the average cluster size for the 10% clustering process and use it as the

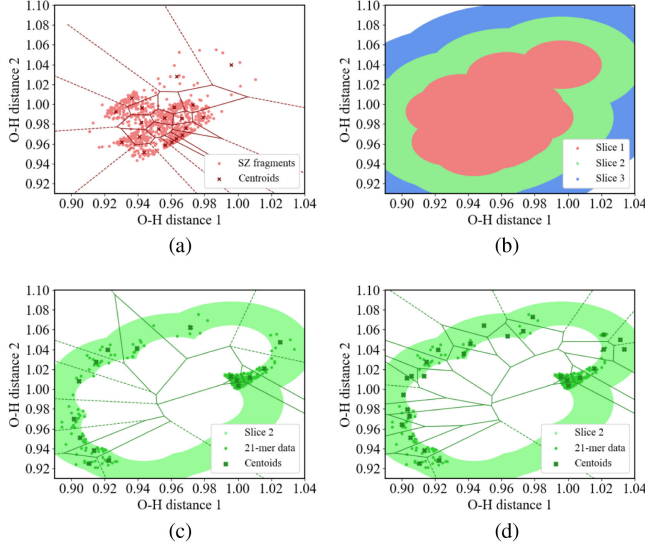


FIG. 31. The process of recursive minibatch- k -means for additional slices partitioning. (a) Slice 1 partitioning. (b) Additional slices in the space. (c) Round 1 coarse partitioning of slice 2. (d) Final partitioning of slice 2.

target for additional slice clustering. The average inertia is computed as

$$\eta_0 = \frac{1}{M} \sum_a^k \sum_{\mathbf{r}_i \in C_j} |\mathbf{r}_i - \bar{\mathbf{r}}_j|^2, \quad (\text{F1})$$

where M is the number of primitive data points used for clustering the fragments found in the solvated Zundel trajectory, k is the number of training data points and also represents the number of k -means clusters or regions into which the M data points are divided, and C_j are the (irregular and multidimensional) regions and also set of points within the j th cluster with $\bar{\mathbf{r}}_j$ being the centroid of the region C_j .

For example, for the H_2O fragment from within the solvated Zundel data as described in Fig. 31(a), η_0 is computed from the red lines partitioning the space, with the $\bar{m}_a$ geometries represented as “X.” We aim to create similar tessellations for additional slices like those within the green area and blue area in Fig. 31(b). We begin with a coarse partitioning of data by using a minibatch- k -means clustering with the number of clusters chosen as the square root of the number of samples in the slice as shown in Fig. 31(c). We label this clustering process as round 1 ($j = 1$) and compute the clusterwise average inertia similar to Eq. (F1) but only for data within the s th slice, l th cluster, iteration round number j as

$$\eta_{s,j,l} = \frac{1}{M_{s,j,l}} \sum_{\mathbf{r}_i \in C_{s,j,l}} |\mathbf{r}_i - \bar{\mathbf{r}}_{s,j,l}|^2, \quad (\text{F2})$$

where $M_{s,j,l}$ is the number of sample points in the s th slice, l th cluster, for iteration round number j . As shown in

Fig. 31(c), some clusters are too large, such as the top and rightmost. The value of $\eta_{s,j,l}$ for every cluster is used to measure the size of the cluster and compare it with the target average inertia of η_0 . If $\eta_{s,j,l} > 2 * \eta_0$, we perform a second round ($j = 2$) of minibatch- k -means clustering on the data within the s th slice, l th cluster, requiring that the l th cluster be further subdivided into the following number of the cluster:

$$k_{s,j+1,l} = \frac{\eta_{s,j,l}}{\eta_0}. \quad (\text{F3})$$

Then, we compute clusterwise average inertia from Eq. (F2) for the round 2 clusters ($j = 2$) and compare their $\eta_{s,j,l}$ with $2 * \eta_0$ again. This recursive process continues until all $\eta_{s,j,l}$ for clusters are below the threshold $2 * \eta_0$ and ultimately lead to space partitioning shown in Fig. 31(d). Here, we can clearly see that the large cluster in Fig. 31(c) such as the rightmost one is further divided into four clusters.

APPENDIX G: THE COMPLEMENTARY FIGURES FOR ALL OTHER FRAGMENTS AS DISCUSSED IN SEC. VIA

In this section, we provide complementary figures for fragments species not covered in the main discussion in Sec. VIA. Specifically, we provide the same set of figures

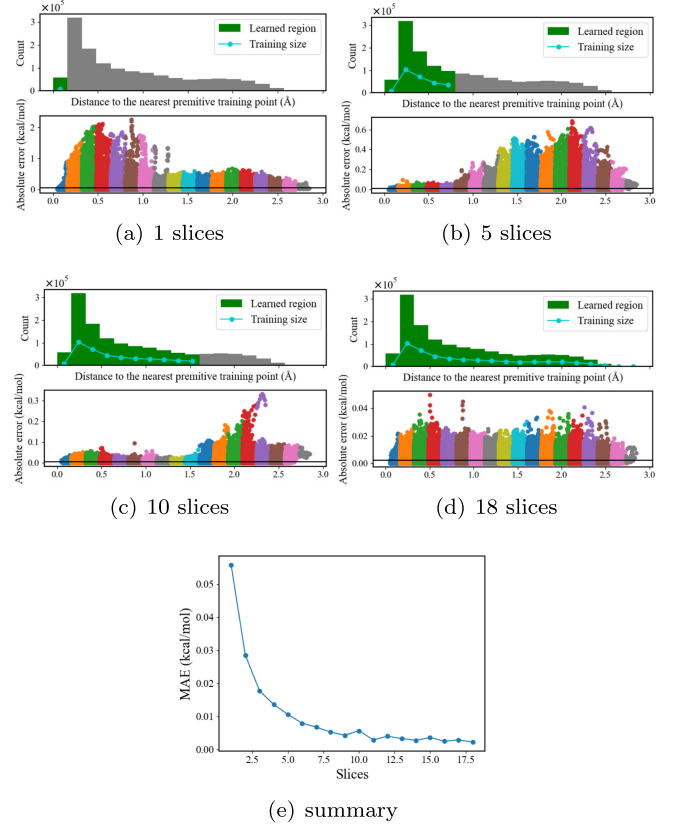


FIG. 32. Error distribution change with number of slices trained for H_4O_2 .

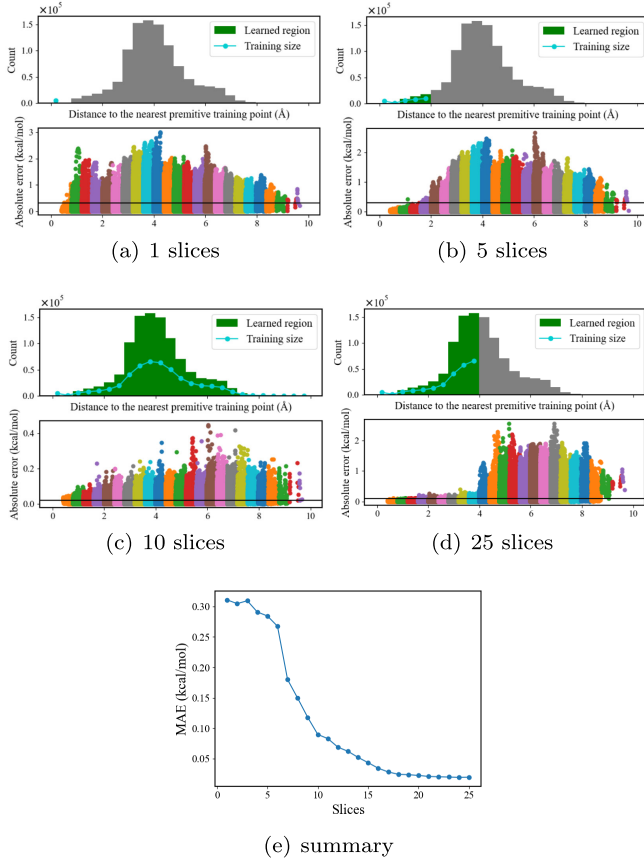


FIG. 33. Error distribution change with number of slices trained for H_6O_3 .

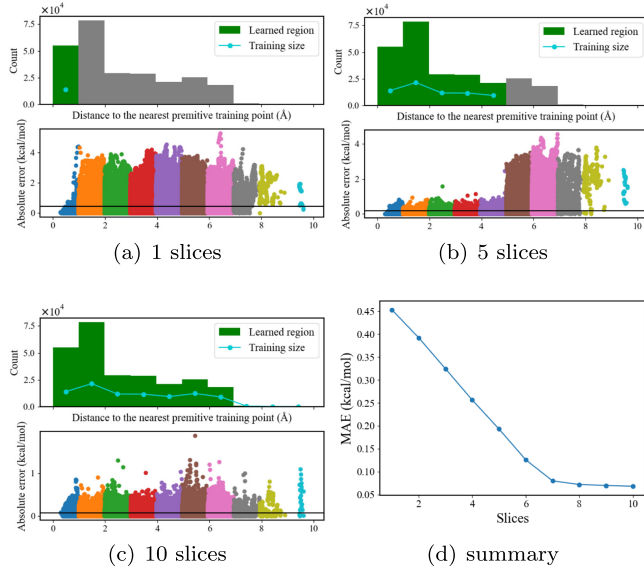


FIG. 34. Error distribution change with number of slices trained for H_7O_3^+ .

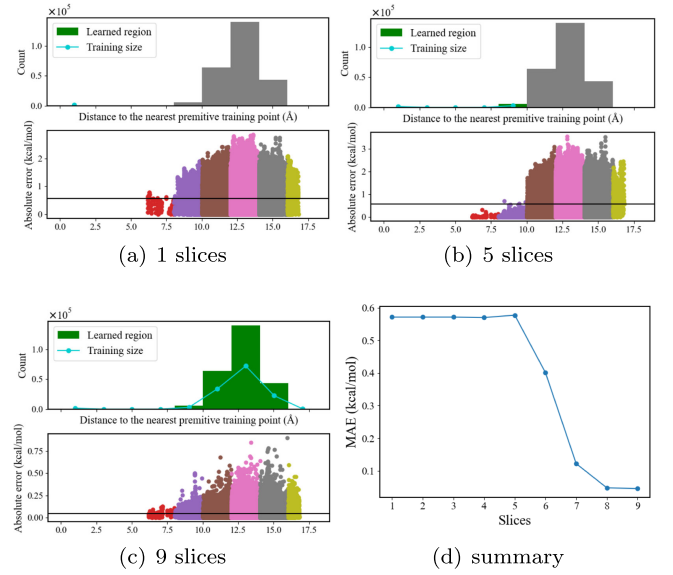


FIG. 35. Error distribution change with number of slices trained for H_8O_4 .

as Fig. 19 for H_4O_2 , H_6O_3 , H_7O_3^+ , H_8O_4 , and H_9O_4^+ , shown in Figs. 32–36, respectively. Overall, these fragments behave similar to that of H_5O_2^+ as discussed in Sec. VI A. Throughout the transferring process, neural networks capture the energy pattern within each additional slice and monotonically reduce prediction errors. As a result, the final predictions are accurate for fragments obtained from both solvated Zundel and protonated 21-water cluster trajectories.

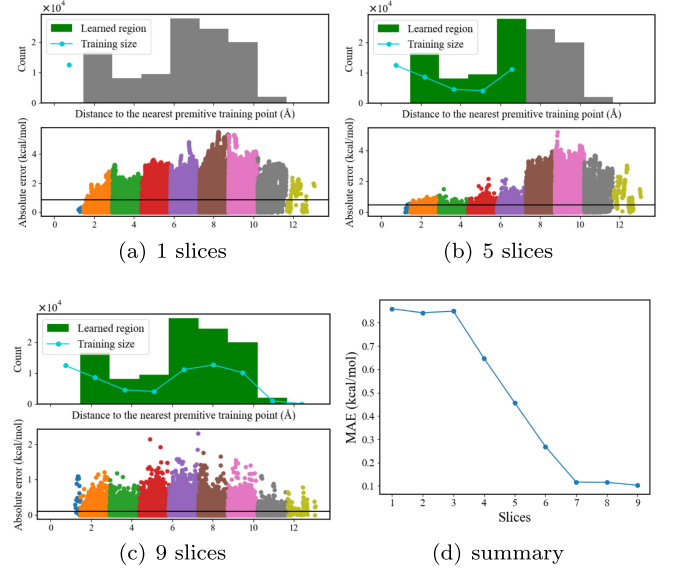


FIG. 36. Error distribution change with number of slices trained for H_9O_4^+ .

- [1] J. Murrell, S. Carter, S. Farantos, P. Huxley, and A. Varandas, *Molecular Potential Energy Functions* (Wiley, New York, 1984).
- [2] B. J. Braams and J. M. Bowman, *Permutationally invariant potential energy surfaces in high dimensionality*, *Int. Rev. Phys. Chem.* **28**, 577 (2009).
- [3] Z. Xie and J. M. Bowman, *Permutationally invariant polynomial basis for molecular energy surface fitting via monomial symmetrization*, *J. Chem. Theory Comput.* **6**, 26 (2010).
- [4] Ö. F. Alış and H. Rabitz, *Efficient implementation of high dimensional model representations*, *J. Math. Chem.* **29**, 127 (2001).
- [5] A. Jäckle and H.-D. Meyer, *Product representation of potential energy surfaces*, *J. Chem. Phys.* **104**, 7974 (1996).
- [6] D. Peláez and H.-D. Meyer, *The multigrid POTFIT (MGPF) method: Grid representations of potentials for quantum dynamics of large systems*, *J. Chem. Phys.* **138**, 014108 (2013).
- [7] A. Leclerc and T. Carrington, *Calculating vibrational spectra with sum of product basis functions without storing full-dimensional vectors or matrices*, *J. Chem. Phys.* **140**, 174111 (2014).
- [8] C. Qu, R. Conte, P. L. Houston, and J. M. Bowman, *Full-dimensional potential energy surface for acetylacetone and tunneling splittings*, *Phys. Chem. Chem. Phys.* **23**, 7758 (2021).
- [9] W. Grossmann, S. Eilermann, T. Rensmeyer, A. Liebert, M. Hohmann, C. Wittke, and O. Niggemann, *Position paper on materials design—A modern approach*, [arXiv:2312.10996](https://arxiv.org/abs/2312.10996).
- [10] R. Vasudevan, G. Pilania, and P. V. Balachandran, *Machine learning for materials design and discovery*, *J. Appl. Phys.* **129**, 070401 (2021).
- [11] S. Axelrod, D. Schwalbe-Koda, S. Mohapatra, J. Damewood, K. P. Greenman, and R. Gómez-Bombarelli, *Learning matter: Materials design with machine learning and atomistic simulations*, *Acc. Mater. Res.* **3**, 343 (2022).
- [12] K. Persson, Y. Hinuma, Y. S. Meng, A. Van der Ven, and G. Ceder, *Thermodynamic and kinetic properties of the li-graphite system from first-principles calculations*, *Phys. Rev. B* **82**, 125416 (2010).
- [13] D. Flam-Shepherd, A. Zhigalin, and A. Aspuru-Guzik, *Scalable fragment-based 3D molecular design with reinforcement learning*, [arXiv:2202.00658](https://arxiv.org/abs/2202.00658).
- [14] R. P. Feynman, *Simulating physics with computers*, *Int. J. Theor. Phys.* **21**, 467 (1982).
- [15] R. P. Feynman, J. Hey, and R. W. Allen, *Feynman Lectures on Computation* (Addison-Wesley Longman Publishing Co., Boston, MA, 1998).
- [16] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
- [17] H.-D. Meyer, U. Manthe, and L. S. Cederbaum, *The multi-configurational time-dependent Hartree approach*, *Chem. Phys. Lett.* **165**, 73 (1990).
- [18] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).
- [19] H. Rabitz and Ö. F. Alış, *General foundations of high dimensional model representations*, *J. Math. Chem.* **25**, 197 (1999).
- [20] S. Manzhos and T. Carrington Jr, *Using neural networks, optimized coordinates, and high-dimensional model representations to obtain a vinyl bromide potential surface*, *J. Chem. Phys.* **129**, 224104 (2008).
- [21] F. Otto, *Multi-layer POTFIT: An accurate potential representation for efficient high-dimensional quantum dynamics*, *J. Chem. Phys.* **140**, 014106 (2014).
- [22] I. Sumner and S. S. Iyengar, *Quantum wavepacket ab initio molecular dynamics: An approach for computing dynamically averaged vibrational spectra including critical nuclear quantum effects*, *J. Phys. Chem. A* **111**, 10313 (2007).
- [23] N. DeGregorio and S. S. Iyengar, *Efficient and adaptive methods for computing accurate potential surfaces for quantum nuclear effects: Applications to hydrogen-transfer reactions*, *J. Chem. Theory Comput.* **14**, 30 (2018).
- [24] H. B. Schlegel and M. J. Frisch, *Computational bottlenecks in molecular orbital calculations*, in *Theoretical and Computational Models for Organic Chemistry* (Springer, New York, 1991), pp. 5–33.
- [25] C. Schran, F. Brieuc, and D. Marx, *Transferability of machine learning potentials: Protonated water neural network potential applied to the protonated water hexamer*, *J. Chem. Phys.* **154**, 050001 (2021).
- [26] T. T. Nguyen, E. Szekely, G. Imbalzano, J. Behler, G. Csanyi, M. Ceriotti, A. W. Goetz, and F. Paesani, *Comparison of permutationally invariant polynomials, neural networks, and Gaussian approximation potentials in representing water interactions through many-body expansions*, *J. Chem. Phys.* **148**, 241726 (2018).
- [27] C. Schran, J. Behler, and D. Marx, *Automated fitting of neural network potentials at coupled cluster accuracy: Protonated water clusters as testing ground*, *J. Chem. Theory Comput.* **16**, 88 (2020).
- [28] D. Lu, J. Behler, and J. Li, *Accurate global potential energy surfaces for the $\text{H} + \text{CH}_3\text{OH}$ reaction by neural network fitting with permutation invariance*, *J. Phys. Chem. A* **124**, 5737 (2020).
- [29] D. Lu, J. Qi, M. Yang, J. Behler, H. Song, and J. Li, *Mode specific dynamics in the $\text{H}_2 + \text{SH} \rightarrow \text{H} + \text{H}_2\text{S}$ reaction*, *Phys. Chem. Chem. Phys.* **18**, 29113 (2016).
- [30] J. Li, K. Song, and J. Behler, *A critical comparison of neural network potentials for molecular reaction dynamics with exact permutation symmetry*, *Phys. Chem. Chem. Phys.* **21**, 9672 (2019).
- [31] J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. E. Roitberg, *Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning*, *Nat. Commun.* **10**, 2903 (2019).
- [32] A. Li and H. Guo, *A nine-dimensional ab initio global potential energy surface for the $\text{H}_2\text{O}^+ + \text{H}_2 \rightarrow \text{H}_3\text{O}^+ + \text{H}$ reaction*, *J. Chem. Phys.* **140**, 224313 (2014).
- [33] F. Lu, L. Cheng, R. J. DiRisio, J. M. Finney, M. A. Boyer, P. Moonkaen, J. Sun, S. J. R. Lee, J. E. Deustua, T. F. Miller, III, and A. B. McCoy, *Fast near ab initio potential*

- energy surfaces using machine learning, *J. Phys. Chem. A* **126**, 4013 (2022).
- [34] A. Aldossary, J. A. Campos-Gonzalez-Angulo, S. Pablo-García, S. X. Leong, E. M. Rajaonson, L. Thiede, G. Tom, A. Wang, D. Avagliano, and A. Aspuru-Guzik, *In silico chemical experiments in the age of AI: From quantum chemistry to machine learning and back*, *Adv. Mater.* **36**, 2402369 (2024).
- [35] J. Behler and M. Parrinello, *Generalized neural-network representation of high-dimensional potential-energy surfaces*, *Phys. Rev. Lett.* **98**, 146401 (2007).
- [36] J. Behler, *Four generations of high-dimensional neural network potentials*, *Chem. Rev.* **121**, 10037 (2021).
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, *Attention is all you need*, in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017), Vol. 30.
- [38] J. Li and S. S. Iyengar, *Ab initio molecular dynamics using recursive, spatially separated, overlapping model subsystems mixed within an oniom based fragmentation energy extrapolation technique*, *J. Chem. Theory Comput.* **11**, 3978 (2015).
- [39] J. Li, C. Haycraft, and S. S. Iyengar, *Hybrid, extended lagrangian—Born-Oppenheimer ab initio molecular dynamics using fragment-based electronic structure*, *J. Chem. Theory Comput.* **12**, 2493 (2016).
- [40] C. Haycraft, J. Li, and S. S. Iyengar, *Efficient, “on-the-fly” Born–Oppenheimer and Car–Parrinello–type dynamics with coupled cluster accuracy through fragment based electronic structure*, *J. Chem. Theory Comput.* **13**, 1887 (2017).
- [41] T. C. Ricard, C. Haycraft, and S. S. Iyengar, *Adaptive, geometric networks for efficient coarse-grained ab initio molecular dynamics with Post-Hartree-Fock accuracy*, *J. Chem. Theory Comput.* **14**, 2852 (2018).
- [42] T. C. Ricard and S. S. Iyengar, *Efficiently capturing weak interactions in ab initio molecular dynamics through “on-the-fly” basis set extrapolation*, *J. Chem. Theory Comput.* **14**, 5535 (2018).
- [43] A. Kumar and S. S. Iyengar, *Fragment-based electronic structure for potential energy surfaces using a superposition of fragmentation topologies*, *J. Chem. Theory Comput.* **15**, 5769 (2019).
- [44] T. C. Ricard and S. S. Iyengar, *An efficient and accurate approach to estimate hybrid functional and large basis set contributions to condensed phase systems and molecule-surface interactions*, *J. Chem. Theory Comput.* **16**, 4790 (2020).
- [45] T. C. Ricard, A. Kumar, and S. S. Iyengar, *Embedded, graph-theoretically defined many-body approximations for wavefunction-in-DFT and DFT-in-DFT: Applications to gas- and condensed-phase ab initio molecular dynamics, and potential surfaces for quantum nuclear effects*, *Int. J. Quantum Chem.* **120**, e26244 (2020).
- [46] T. C. Ricard, X. Zhu, and S. S. Iyengar, *Capturing weak interactions in surface adsorbate systems at coupled cluster accuracy: A graph-theoretic molecular fragmentation approach improved through machine learning*, *J. Chem. Theory Comput.* **19**, 8541 (2023).
- [47] A. Kumar, N. DeGregorio, and S. S. Iyengar, *Graph-theory-based molecular fragmentation for efficient and accurate potential surface calculations in multiple dimensions*, *J. Chem. Theory Comput.* **17**, 6671 (2021).
- [48] J. H. Zhang, T. C. Ricard, C. Haycraft, and S. S. Iyengar, *Weighted-graph-theoretic methods for many-body corrections within oniom: Smooth aimd and the role of high-order many-body terms*, *J. Chem. Theory Comput.* **17**, 2672 (2021).
- [49] X. Zhu and S. S. Iyengar, *Graph theoretic molecular fragmentation for multidimensional potential energy surfaces yield an adaptive and general transfer machine learning protocol*, *J. Chem. Theory Comput.* **18**, 5125 (2022).
- [50] J. H. Zhang and S. S. Iyengar, *Graph-|Q⟩⟨C|: A graph-based quantum-classical algorithm for efficient electronic structure on hybrid quantum/classical hardware systems: Improved quantum circuit depth performance*, *J. Chem. Theory Comput.* **18**, 2885 (2022).
- [51] A. Kumar, N. DeGregorio, T. Ricard, and S. S. Iyengar, *Graph-theoretic molecular fragmentation for potential surfaces leads naturally to a tensor network form and allows accurate and efficient quantum nuclear dynamics*, *J. Chem. Theory Comput.* **18**, 7243 (2022).
- [52] S. S. Iyengar, D. Saha, A. Dwivedi, M. A. Lopez-Ruiz, A. Kumar, J. H. Zhang, T. C. Ricard, P. Richerme, and A. Sabry, *Quantum algorithms for the study of electronic structure and molecular dynamics: Novel computational protocols*, in *Comprehensive Computational Chemistry* (Elsevier, New York, 2023).
- [53] S. S. Iyengar, J. H. Zhang, D. Saha, and T. C. Ricard, *Graph-|Q⟩⟨C|: A quantum algorithm with reduced quantum circuit depth for electronic structure*, *J. Phys. Chem. A* **127**, 9334 (2023).
- [54] S. S. Iyengar, T. C. Ricard, and X. Zhu, *A reformulation of all oniom-type molecular fragmentation approaches using graph theory-based projection operators: Applications to dynamics, molecular potential surfaces, and machine learning and quantum computing*, *J. Phys. Chem. A* **128**, 466 (2024).
- [55] W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka, *Parallel distributed processing model with local space-invariant interconnections and its optical architecture*, *Appl. Opt.* **29**, 4790 (1990).
- [56] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, *Convolutional networks on graphs for learning molecular fingerprints*, in *Proceedings of the 29th International Conference on Neural Information Processing Systems, NIPS’15* (MIT Press, Cambridge, MA, 2015), pp. 2224–2232.
- [57] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, *Graph attention networks*, *arXiv*: 1710.10903.
- [58] J. A. Fournier, C. T. Wolke, M. A. Johnson, T. T. Obadrakh, K. D. Jordan, S. M. Kathmann, and S. S. Xantheas, *Snapshots of proton accommodation at a microscopic water surface: Understanding the vibrational*

- spectral signatures of the charge defect in cryogenically cooled $\text{H}(\text{H}_2\text{O})_n^+ = 2 - 28$ clusters, *J. Phys. Chem. A* **119**, 37 (2015).
- [59] J.-W. Shin, N. I. Hammer, E. G. Diken, M. A. Johnson, R. S. Walters, T. D. Jaeger, M. A. Duncan, R. A. Christie, and K. D. Jordan, *Infrared signature of structures associated with the $\text{H}^+(\text{H}_2\text{O})_n$ ($n = 6$ to 27) clusters*, *Science* **304**, 1137 (2004).
- [60] S. S. Iyengar, M. K. Petersen, T. J. F. Day, C. J. Burnham, V. E. Teige, and G. A. Voth, *The properties of ion-water clusters. I. The protonated 21-water cluster*, *J. Chem. Phys.* **123**, 084309 (2005).
- [61] S. S. Iyengar, *Further analysis of the dynamically averaged vibrational spectrum for the “magic” protonated 21-water cluster*, *J. Chem. Phys.* **126**, 216101 (2007).
- [62] N. I. Hammer, E. G. Diken, J. R. Roscioli, M. A. Johnson, E. M. Myshakin, K. D. Jordan, A. B. McCoy, X. Huang, J. M. Bowman, and S. Carter, *The vibrational predissociation spectra of the $\text{H}_5\text{O}_2^+ \cdot \text{RG}_n$ ($\text{RG} = \text{Ar}, \text{Ne}$) clusters: Correlation of the solvent perturbations in the free oh and shared proton transitions of the Zundel ion*, *J. Chem. Phys.* **122**, 244301 (2005).
- [63] J. M. Headrick, E. G. Diken, R. S. Walters, N. I. Hammer, R. A. Christie, J. Cui, E. M. Myshakin, M. A. Duncan, M. A. Johnson, and K. Jordan, *Spectral signatures of hydrated proton vibrations in water clusters*, *Science* **308**, 1765 (2005).
- [64] S. M. Dietrick and S. S. Iyengar, *Constructing periodic phase space orbits from ab initio molecular dynamics trajectories to analyze vibrational spectra: Case study of the Zundel (H_5O_2^+) cation*, *J. Chem. Theory Comput.* **8**, 4876 (2012).
- [65] K. R. Asmis, N. L. Pivonka, G. Santambrogio, M. Brümmer, C. Kaposta, D. M. Neumark, and L. Wöste, *Gas-phase infrared spectrum of the protonated water dimer*, *Science* **299**, 1375 (2003).
- [66] O. Vendrell, F. Gatti, and H.-D. Meyer, *Dynamics and infrared spectroscopy of the protonated water dimer*, *Angew. Chem., Int. Ed.* **46**, 6918 (2007).
- [67] E. G. Diken, J. M. Headrick, J. R. Roscioli, J. C. Bopp, M. A. Johnson, and A. B. McCoy, *Fundamental excitations of the shared proton in the H_3O_2^- and H_5O_2^+ complexes*, *J. Phys. Chem. A* **109**, 1487 (2005).
- [68] W. H. Robertson, E. G. Diken, E. A. Price, J.-W. Shin, and M. A. Johnson, *Spectroscopic determination of the oh-solvation shell in the $\text{OH}^-(\text{H}_2\text{O})_n$ clusters*, *Science* **299**, 1367 (2003).
- [69] G. Lorentz, M. Golitschek, and Y. Makovoz, *Constructive Approximation* (Springer, New York, 1996).
- [70] F. Girosi and T. Poggio, *Representation properties of networks: Kolmogorov’s theorem is irrelevant*, *Neural Comput.* **1**, 465 (1989).
- [71] I. M. Sobol, *On the distribution of points in a cube and the approximate evaluation of integrals*, *USSR Comput. Math. Math. Phys.* **7**, 86 (1967).
- [72] B. Roux, *Computational studies of the gramicidin channel*, *Acc. Chem. Res.* **35**, 366 (2002).
- [73] T. W. Allen, T. Baştuğ, S. Kuyucak, and S.-H. Chung, *Gramicidin A channel as a test ground for molecular dynamics force fields*, *Biophys. J.* **84**, 2159 (2003).
- [74] A. Pullman and C. Etchebest, *The gramicidin A channel: The energy profile for single and double occupancy in a head-to-head β_6 . 33, 3-helical dimer backbone*, *FEBS Lett.* **163**, 199 (1983).
- [75] A. Björklund, T. Husfeldt, and M. Koivisto, *Set partitioning via inclusion exclusion*, *SIAM J. Comput.* **39**, 546 (2009).
- [76] T. K. Dey and N. R. Shah, *On the number of simplicial complexes in $\mathbb{R}^d$* , *Comput. Geom.* **8**, 267 (1997).
- [77] C. C. Adams and R. D. Franzosa, *Introduction to Topology: Pure and Applied* (Pearson Prentice Hall, Harlow, UK, 2008).
- [78] M. Berger, P. Pansu, J.-P. Berry, and X. Saint-Raymond, *Affine spaces*, in *Problems in Geometry* (Springer, New York, 1984), p. 11.
- [79] A. Bowyer, *Computing Dirichlet tessellations*, *Comput. J.* **24**, 162 (1981).
- [80] D. Watson, *Computing the n-dimensional Delaunay tessellation with applications to Voronoi polytopes*, *Comput. J.* **24**, 167 (1981).
- [81] F. Aurenhammer, *Voronoi diagrams—A survey of a fundamental geometric data structure*, *ACM Comput. Surv.* **23**, 345 (1991).
- [82] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations—Concepts and Applications of Voronoi Diagrams* (John Wiley and Sons, New York, 2000).
- [83] S. Hert and M. Seel, *CGAL User and Reference Manual*, 4th ed. (CGAL Editorial Board, 2018), Chap. dD Convex Hulls and Delaunay Triangulations.
- [84] G. Farin, *Surfaces over Dirichlet tessellations*, *Comput. Aided Geom. Des.* **7**, 281 (1990).
- [85] F. Maseras and K. Morokuma, *Imomm: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states*, *J. Comput. Chem.* **16**, 1170 (1995).
- [86] A. Nandi, C. Qu, P. L. Houston, R. Conte, and J. M. Bowman, *Δ -machine learning for potential energy surfaces: A PIP approach to bring a DFT-based PES to CCSD(T) level of theory*, *J. Chem. Phys.* **154**, 051102 (2021).
- [87] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, *Big data meets quantum chemistry approximations: The Δ -machine learning approach*, *J. Chem. Theory Comput.* **11**, 2087 (2015).
- [88] J. Chen, W. Xu, and R. Zhang, *Δ -machine learning-driven discovery of double hybrid organic–inorganic perovskites*, *J. Mater. Chem. A* **10**, 1402 (2022).
- [89] P. A. Unzueta, C. S. Greenwell, and G. J. O. Beran, *Predicting density functional theory-quality nuclear magnetic resonance chemical shifts via Δ -machine learning*, *J. Chem. Theory Comput.* **17**, 826 (2021).
- [90] A. Nandi, P. Pandey, P. L. Houston, C. Qu, Q. Yu, R. Conte, A. Tkatchenko, and J. M. Bowman, *Δ -machine learning to elevate DFT-based potentials and a force field to the CCSD(T) level illustrated for ethanol*, *J. Chem. Theory Comput.* **20**, 8807 (2024).
- [91] F. Riesz and B. Sz. Nagy, *Functional Analysis* (Dover, New York, 1990).
- [92] D. Bahdanau, K. Cho, and Y. Bengio, *Neural machine translation by jointly learning to align and translate*, *arXiv:1409.0473*.

- [93] G. Zundel, *Hydration and Intermolecular Interaction: Infrared Investigations with Polyelectrolyte Membranes* (Academic Press, New York, 2012).
- [94] M. E. Tuckerman, P. J. Ungar, T. Vonrosenvinge, and M. L. Klein, *Ab initio molecular dynamics simulations*, *J. Phys. Chem.* **100**, 12878 (1996).
- [95] M. Tuckerman, K. Laasonen, M. Sprik, and M. Parrinello, *J. Phys. Chem.* **99**, 5749 (1995).
- [96] U. W. Schmitt and G. A. Voth, *The computer simulation of proton transport in water*, *J. Chem. Phys.* **111**, 9361 (1999).
- [97] U. W. Schmitt and G. A. Voth, *Multistate empirical valence bond model for proton transport in water*, *J. Phys. Chem. B* **102**, 5547 (1998).
- [98] R. Pomes and B. Roux, *Structure and dynamics of a proton wire: A theoretical study of H^+ translocation along the single-file water chain in the gramicidin a channel*, *Biophys. J.* **71**, 19 (1996).
- [99] M. J. McEwan and L. F. Phillips, *Chemistry of the Atmosphere* (Edward Arnold, London, 1975).
- [100] R. P. Wayne, *Chemistry of the Atmosphere* (Clarendon Press, Oxford, 1994).
- [101] R. Pomes and B. Roux, *Theoretical study of H^+ translocation along a model proton wire*, *J. Phys. Chem.* **100**, 2519 (1996).
- [102] H. Decornez, K. Drukker, and S. Hammes-Schiffer, *Solvation and hydrogen-bonding effects on proton wires*, *J. Phys. Chem. A* **103**, 2891 (1999).
- [103] M. L. Brewer, U. W. Schmitt, and G. A. Voth, *The formation and dynamics of proton wires in channel environments*, *Biophys. J.* **80**, 1691 (2001).
- [104] M. Teeter, *Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6014 (1984).
- [105] S. Neidle, H. M. Berman, and H. S. Shieh, *Highly structured water network in crystals of a deoxydinucleoside-drug complex*, *Nature (London)* **288**, 129 (1980).
- [106] L. A. Lipscomb, M. E. Peek, F. X. Zhou, J. A. Bertrand, D. VanDerveer, and L. D. Williams, *Water ring structure at DNA interfaces: Hydration and dynamics of DNA-anthracene complexes*, *Biochemistry* **33**, 3649 (1994).
- [107] C. Tu, R. S. Rowlett, B. C. Tripp, J. G. Ferry, and D. N. Silverman, *Chemical rescue of proton transfer in catalysis by carbonic anhydrases in the β - and γ -class*, *Biochemistry* **41**, 15429 (2002).
- [108] S. M. Dietrick and S. S. Iyengar, *Constructing periodic phase space orbits from ab initio molecular dynamics trajectories to analyze vibrational spectra: Case study of the Zundel ($H_5O_2^+$) cation*, *J. Chem. Theory Comput.* **8**, 4876 (2012).
- [109] S. Sadhukhan, D. Munoz, C. Adamo, and G. E. Scuseria, *Predicting proton transfer barriers with density functional methods*, *Chem. Phys. Lett.* **306**, 83 (1999).
- [110] Y. Xie, R. B. Remington, and H. F. Schaefer, III, *The protonated water dimer: Extensive theoretical studies of $H_5O_2^+$* , *J. Chem. Phys.* **101**, 4878 (1994).
- [111] J. Klimes and A. Michaelides, *Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory*, *J. Chem. Phys.* **137**, 120901 (2012).
- [112] R. Peverati and D. Truhlar, *The quest for a universal density functional: The accuracy of density functionals across a broad spectrum of databases in chemistry and physics*, *Phil. Trans. R. Soc. A* **372**, 20120476 (2014).
- [113] A. J. Cohen, P. Mori-Sánchez, and W. Yang, *Challenges for density functional theory*, *Chem. Rev.* **112**, 289 (2012).
- [114] J. A. Hartigan and M. A. Wong, *Algorithm as 136: A k-means clustering algorithm*, *J. R. Stat. Soc. C* **28**, 100 (1979).
- [115] Microsoft, Copilot (July 10 version) (2025), large language model.
- [116] Google Gemini, Gemini AI Chatbot (2025) (accessed: 2025-07-10). Version [Gemini Advanced].
- [117] H. B. Schlegel and M. J. Frisch, *Computational bottlenecks in molecular orbital calculations*, in *Theoretical and Computational Models for Organic Chemistry*, edited by S. J. Formosinho, I. G. Csizmadia, and L. G. Arnaut (Springer, Netherlands, Dordrecht, 1991), pp. 5–33.
- [118] L.-H. Lim and B. J. Nelson, *What is an equivariant neural network?*, *arXiv:2205.07362*.
- [119] R. Kondor, *The principles behind equivariant neural networks for physics and chemistry*, *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2415656122 (2025).
- [120] G. Zundel, *Hydration and Intermolecular Interaction* (Academic Press, New York, 1969).