

# Minimizing Redundancy and Data Requirements of Machine Learning Potential: A Case Study in Interface Combustion

Xiaoya Chang, Di Zhang, Qingzhao Chu,\* and Dongping Chen\*



Cite This: *J. Chem. Theory Comput.* 2024, 20, 6813–6825



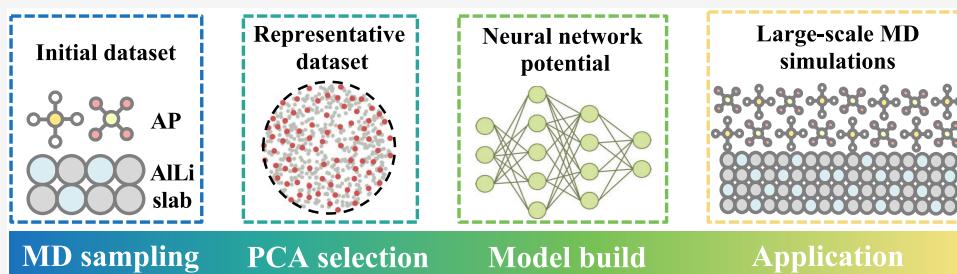
Read Online

ACCESS |

Metrics & More

Article Recommendations

Supporting Information



**ABSTRACT:** The machine learning potential has emerged as a promising approach for addressing the accuracy-versus-efficiency dilemma in molecular modeling. Efficiently exploring chemical spaces with high accuracy presents a significant challenge, particularly for the interface reaction system. This study introduces a workflow aimed at achieving this goal by incorporating the classical SOAP descriptor and practical PCA strategy to minimize redundancy and data requirements, while successfully capturing the features of complex potential energy surfaces. Specifically, the study focuses on interface combustion behaviors within promising alloy-based solid propellants. A neural network potential model tailored for modeling AlLi–AP interface reactions under varying conditions is constructed, showcasing excellent predictive capabilities in energy prediction, force estimation, and bond energies. A series of large-scale MD simulations reveal that Li doping significantly influences the initial combustion stage, enhancing reactivity and reducing thermal conductivity. Mass transfer analysis also highlights the considerably higher diffusion coefficient of Li compared to Al, with the former being three times greater. Consequently, the overall combustion process is accelerated by approximately 10%. These breakthroughs pave the way for virtual screening and the rational design of advanced propellant formulations and microstructures incorporating alloy-formula propellants.

## 1. INTRODUCTION

Molecular modeling is a potent and practical tool employed in the exploration of materials science,<sup>1,2</sup> biological macromolecules,<sup>3,4</sup> and chemical reactions<sup>5,6</sup> at a microscopic level. Computational simulation methods commonly used include precise yet computationally intense *ab initio* approaches, as well as fast classical force field-based methods that may have limitations in applicability and accuracy.<sup>7</sup> Machine learning has emerged as a promising approach to tackle the accuracy-versus-efficiency dilemma in molecular simulations, enabling the exploration of chemical spaces with remarkable efficiency and good accuracy.<sup>8–11</sup> Unlike traditional methods that rely on solving equations derived from physical laws governing the structure–property relationship, machine learning focuses on establishing the functional relationship between chemical descriptors and molecular properties, extending it to unfamiliar systems.<sup>12–14</sup> Representative algorithms for constructing the chemical potential energy surface (PES) include kernel ridge regression (KRR), support vector regression (SVR), and multilayer neural networks (NNs). Among these, state-of-the-art NN-based methods stand out as the most promising, as

they have the potential to approximate any real-valued function with desirable precision.<sup>15,16</sup>

The quality and size of the training data set play a crucial role in fully harnessing the potential of machine learning, significantly affecting the productivity and generalizability of developed models.<sup>17,18</sup> Many studies emphasize the importance of gathering extensive data and strive to cover the PES in the target scenario, which often results in a considerable computational cost. To circumvent this issue, great emphasis is placed on refining the information and reducing redundancy within the data set, focusing on selecting representative configurations rather than numerous similar ones. Behler and colleagues<sup>19</sup> implemented training on multiple NN models to assess a wider range of trial configurations. When a specific

Received: May 2, 2024

Revised: July 13, 2024

Accepted: July 15, 2024

Published: July 29, 2024



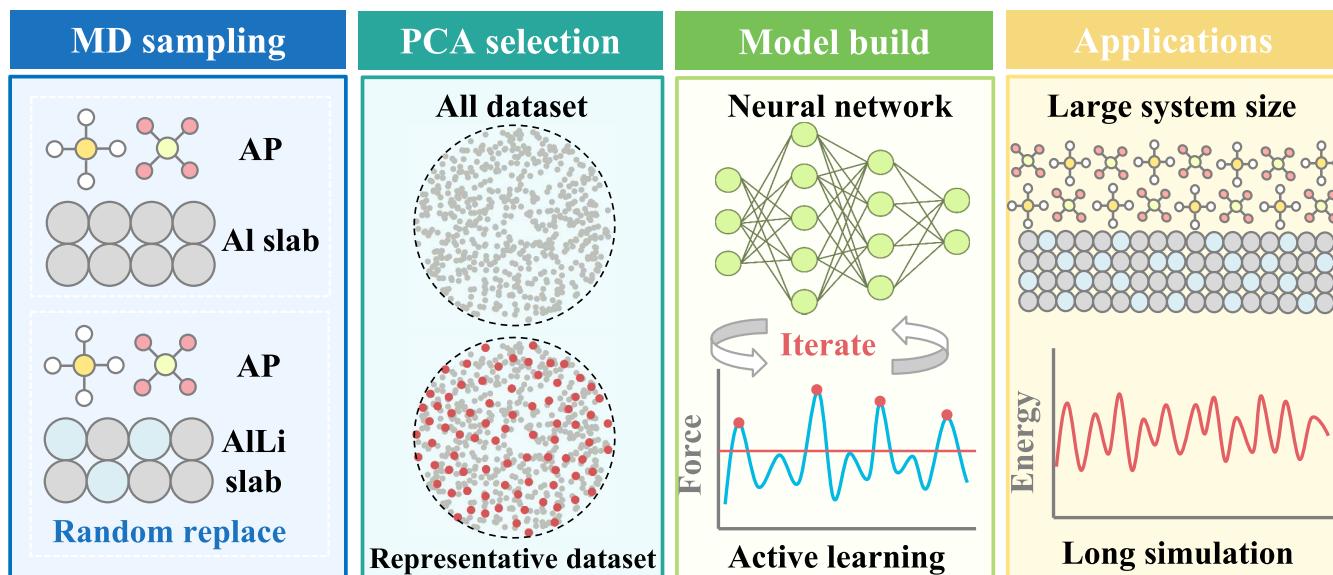


Figure 1. Workflow for data set construction.

structure varies significantly from the training data used for the models, it is expected that the predictions produced by these models will exhibit notable inconsistencies. This approach enabled the identification of underrepresented regions within the configuration space, a concept commonly referred to as the “active learning” or “learning on-the-fly” algorithm. This approach has been implemented in numerous studies. For instance, Zhang et al.<sup>20</sup> established the deep potential generator (DPGen), and Roitberg and co-workers<sup>21</sup> developed the active learning-based ANI-1x potential. These attempts, featured as “learning from mistakes”, are not limited to energy or force prediction, but have been further expanded, such as employing certainty of model instability to guide sampling,<sup>22</sup> incorporating informative configurations with the maximum mean absolute error into the training set.<sup>23</sup> In addition to the workflows mentioned above that are focused on predicting discrepancies, increased efforts<sup>24–26</sup> have been directed toward comparing the chemical descriptors of candidate configurations. Fonseca et al.<sup>17</sup> clustered the chemical space into subregions with similar geometries and energetics, and populated the training set of the model with representatives from the most inaccurate regions. Similarly, Imbalzano and co-workers<sup>27</sup> employed various approaches to select fingerprints as diverse as possible, striking a balance between accuracy and computational efficiency, including CUR decomposition, farthest point sampling, and Pearson correlation-based methods. Xu and his group<sup>28</sup> employed Coulomb matrix and symmetry functions to assess the structural variance among molecules in the data set. They also implemented an improved self-organizing incremental NN to automatically generate a data set encompassing the desired chemical space with minimal redundancy. Additionally, chemists and biologists also employ enhanced sampling techniques to capture the configurational space far from the equilibrium state. These methods include the metadynamic sampling method utilizing biased energy to counteract the free energy surface,<sup>29</sup> and virtual reality-assisted sampling to navigate chemical reaction pathways and reaction sites based on intuition and experience.<sup>30,31</sup> Overall, the above-mentioned active learning, structural similarity comparison, and enhanced sampling

approaches primarily concentrate on constructing and supplementing the data set to enhance the applicability of machine learning models.

Interface sampling poses a significant challenge in realizing the machine learning potential, especially in the domains of catalysis and batteries, particularly in complex combustion reactions.<sup>32</sup> Solid propellants are typically composed of metal additives and an oxidizer, *i.e.*, aluminum (Al) and ammonium perchlorate (AP), which account for approximately 20 and 60% of the composition, respectively. Although Al powder boasts high energy density, cost efficiency, and established safety, its practical implementation is impeded by various factors. The Al particles exhibit prolonged ignition delays due to their naturally coated oxide layer.<sup>33</sup> Moreover, particle agglomeration results in incomplete chemical energy conversion of the propellant and up to a 10% reduction in specific impulse due to two-phase flow losses.<sup>34</sup> To fully exploit its thermal potential, a promising approach is to replace Al powder with its alloy counterparts, such as AlLi alloy.<sup>35</sup> Such a strategy is dedicated to weakening the native oxide layer that forms on the surface of Al powders, thereby enhancing the reactivity of Al-based fuels. More importantly, the considerable difference in boiling points between Li (1620–1643 K<sup>36</sup>) and Al (2740 K<sup>37</sup>) facilitates intraparticle boiling within the AlLi alloy droplets. This phenomenon enhances the effective fragmentation and atomization of the droplets, leading to microexplosions during combustion.<sup>35</sup> In addition, the introduction of Li into AP composite propellant formulations has demonstrated a significant reduction of over 95% in the production of dangerous pollutants, *i.e.*, hydrogen chloride (HCl), and alleviates adverse impacts on the atmospheric ozone layer.<sup>38,39</sup> The combustion process in solid propellants entails chemical reactions among their components. Thermal transfer, mass diffusion, and associated reaction mechanisms at the interface, particularly at the oxidizer and metal additive interface, are paramount in this process. Various experimental efforts have been made to investigate the combustion performance of promising Al–Li alloys with AP, but the comprehensive reaction mechanism remains unresolved. There is an urgent need for microscopic insights into this area.

**Table 1.** Sampling Method and Data Set Description

stage	molecule	ensemble	pressure (MPa)	temperature (K)	before PCA <sup>a</sup>	after PCA <sup>b</sup>
MD sampling	Al + AP	NVT/NPT	10, 100, 500, 1000, 2000, 4000	300–4000	73,514	3161 (3417)
	AlLi + AP <sup>c</sup>	NVT/NPT	10, 100, 500, 1000, 2000, 4000	300–4000	17,085	6822 (6939)
active learning	Al + AP	NVT/NPT	10, 100, 500, 1000, 2000, 4000	300–4000	255	
	AlLi + AP	NVT/NPT	10, 100, 500, 1000, 2000, 4000	300–4000	1013	
total					11,251	

<sup>a</sup>The number of configurations sampled by the MD method. It is important to clarify that these configurations were not subjected to high-level DFT calculations but solely served as candidates for PCA selection. <sup>b</sup>The value in the bracket represents the number of selected configurations from the candidate data set using the PCA algorithm, while the reduced value outside the bracket indicates the number of converged configurations calculated using the CP2K package. <sup>c</sup>In the MD sampling stage, the configurations of AlLi–AP are generated by randomly replacing Al with Li in the Al–AP systems at five different mass ratios of Li doping: 3, 5, 10, 15, and 20 wt %. Each mass ratio includes 3417 configurations.

This study aims to develop a set of deep neural network potential (NNP) models to describe the intricate interface interactions using a streamlined and less-redundant data set. A dimensionality reduction algorithm is utilized to select a representative data set that strikes a balance between computational efficiency and accuracy. The accuracy of the NNP model in predicting interatomic interactions between AlLi additives and the AP oxidizer is evaluated by examining energy, force, bond dissociation, and surface adsorption. Direct atomic simulations of Al–AP and AlLi–AP interface combustion are conducted to analyze heat and mass transfer processes and elaborate on the enhancement mechanism of the AlLi alloy in solid propellants.

## 2. METHODOLOGY

The workflow of this study is depicted in Figure 1, comprising MD sampling, PCA selection, NNP model development, active learning, and large-scale simulations. The subsequent sections include the essential details of each step. Further information on the numerical parameters for each part is provided in the Supporting Information.

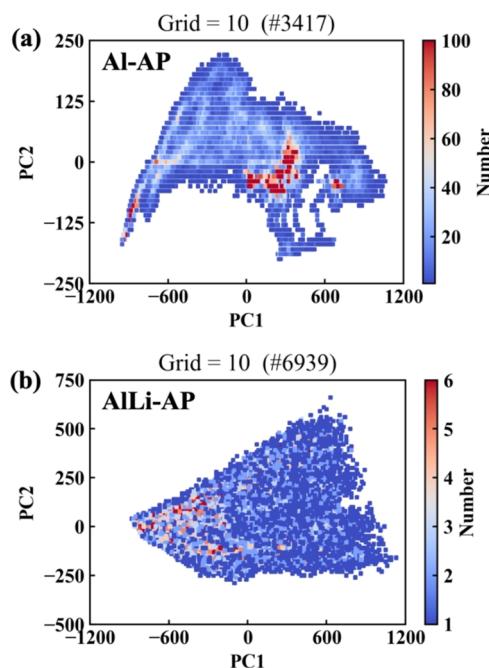
**2.1. MD Sampling.** The initial data set for Al and AP interface combustion was constructed using the published NNP model developed by Chu et al.<sup>40</sup> This set of NNP model takes into account the interface interaction between Al, 1,3,5-trinitroperhydro-1,3,5-triazine (RDX), AP, and hydroxyl-terminated polybutadiene (HTPB) binder with a focus on solid propellant combustion. The proposed model exhibits mean absolute errors (MAE) for energy and atomic force below 40 meV/atom and 0.5 eV/Å, respectively, making it a reliable approach for analyzing the interface interactions in solid propellants. We employ this NNP model and conduct MD simulations on the Al–AP interface to construct the initial data set. Two ensemble simulations are included to explore the configurations on PES, that is, an NVT ensemble heating from 300 to 4000 K, and an NPT ensemble with a pressure range of 10–4000 MPa and temperature range of 300–4000 K. The sampling temperature exceeds both the melting point of Al (933 K<sup>37</sup>) and the decomposition temperature of AP (583–708 K<sup>41</sup>), allowing complete interface combustion. Besides, the internal pressure within the sampling structure (*i.e.*, Al–AP interface) at the NVT ensemble corresponds to ~2000 MPa. The high-pressure conditions (>2000 MPa) align with a compressed region in the curve of the equation of state. Under lower pressure conditions, the reaction products are ejected into the surrounding environment, enabling the simulations of gas-phase combustion. The above MD approach sampled a total of 73,514 configurations for the Al–AP interface reaction, serving as the candidate data set (Table 1) before further

selection. Undertaking high-precision DFT calculations for the entire training set without eliminating redundancy and similarity in the data set incurs considerable computational costs.

**2.2. PCA Selection.** To identify the structural diversity within the initial data set, we utilized the smooth overlap of atomic positions (SOAP) fingerprinting method<sup>42</sup> as implemented in the Dscribe Python library.<sup>43</sup> SOAP encodes the atomic neighborhood surrounding a spatial point or atom by utilizing a Gaussian smeared atomic density expansion with orthonormal basis functions consisting of spherical harmonics and radial basis functions.<sup>44,45</sup> This results in a SOAP fingerprint vector for each atom, which was averaged using the “inner” averaging scheme to obtain a high-dimensional configuration fingerprint for each structure.

In the realm of available dimensionality reduction algorithms, principal component analysis (PCA) has been proven effective for handling high-dimensional SOAP data.<sup>46</sup> The PCA approach is employed to project the structural features of the SOAP descriptor onto a two-dimensional plane. The contributions of these two principal components can be evaluated by their corresponding variation ratio. The obtained PCA plane is then segmented into specified-sized grids according to its distributions and values, *e.g.*, a grid size of 10. This approach assumes that the configurations within the same grid share similar structural characteristics, thereby effectively reducing redundancy among the sampled data set. One configuration is randomly selected from each grid to serve as the representative data set, resulting in a total of 3417 configurations representing the PES of Al–AP, as illustrated in Figure 2a. For the AlLi–AP interface, Li is randomly introduced to replace Al in the selected Al–AP data set, as described in Table 1. To account for practical application, five mass ratios (3, 5, 10, 15, and 20 wt %) were included. Therefore, the AlLi–AP data set initially encompasses 17,085 structures, covering varied configurational space. The PCA workflow is applied to construct the AlLi–AP data set with a grid size of 10, resulting in a total of 6939 configurations in Figure 2b. Only the selected data sets from these two sets (*e.g.*, Al–AP and AlLi–AP) undergo high-level DFT calculations. Among these calculations, a total of 9983 configurations are converged and serve as the initial candidates for training the NNP model before active learning.

**2.3. NNP Model and Active Learning.** This study employs the Deep Potential (DP) scheme, developed by Zhang and colleagues,<sup>47</sup> which has gained increasing popularity in the fields of chemistry and materials research.<sup>48,49</sup> This innovative technique shows promising potential for both finite and extended molecular systems, as well as in metallic and



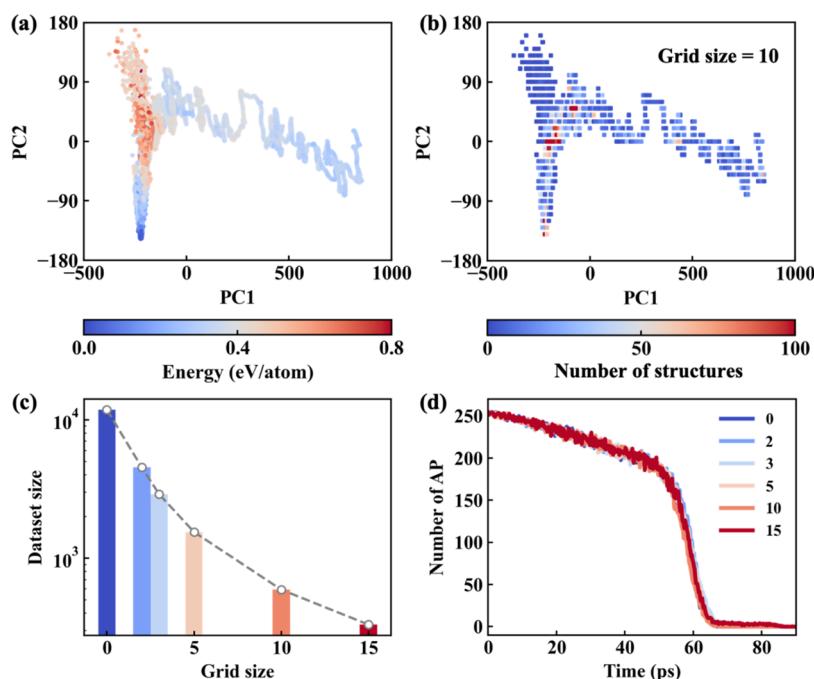
**Figure 2.** PCA plane for the (a) Al-AP interface and (b) AlLi-AP interface partitioned into grids with a size of 10 units. The color map represents the number of structures within each grid.

chemically bonded systems.<sup>50</sup> In the DP framework, the PES is represented by a deep NN model with three hidden layers, enabling the capture of intricate features on the PES. The atomic coordination ( $R$ ) is interpreted into systematic energies ( $E$ ) and atomic forces ( $F$ ) through a chemical environment descriptor, e.g., DeepPot-Se descriptor, ensuring rotation and translation invariance of the structure. More information about

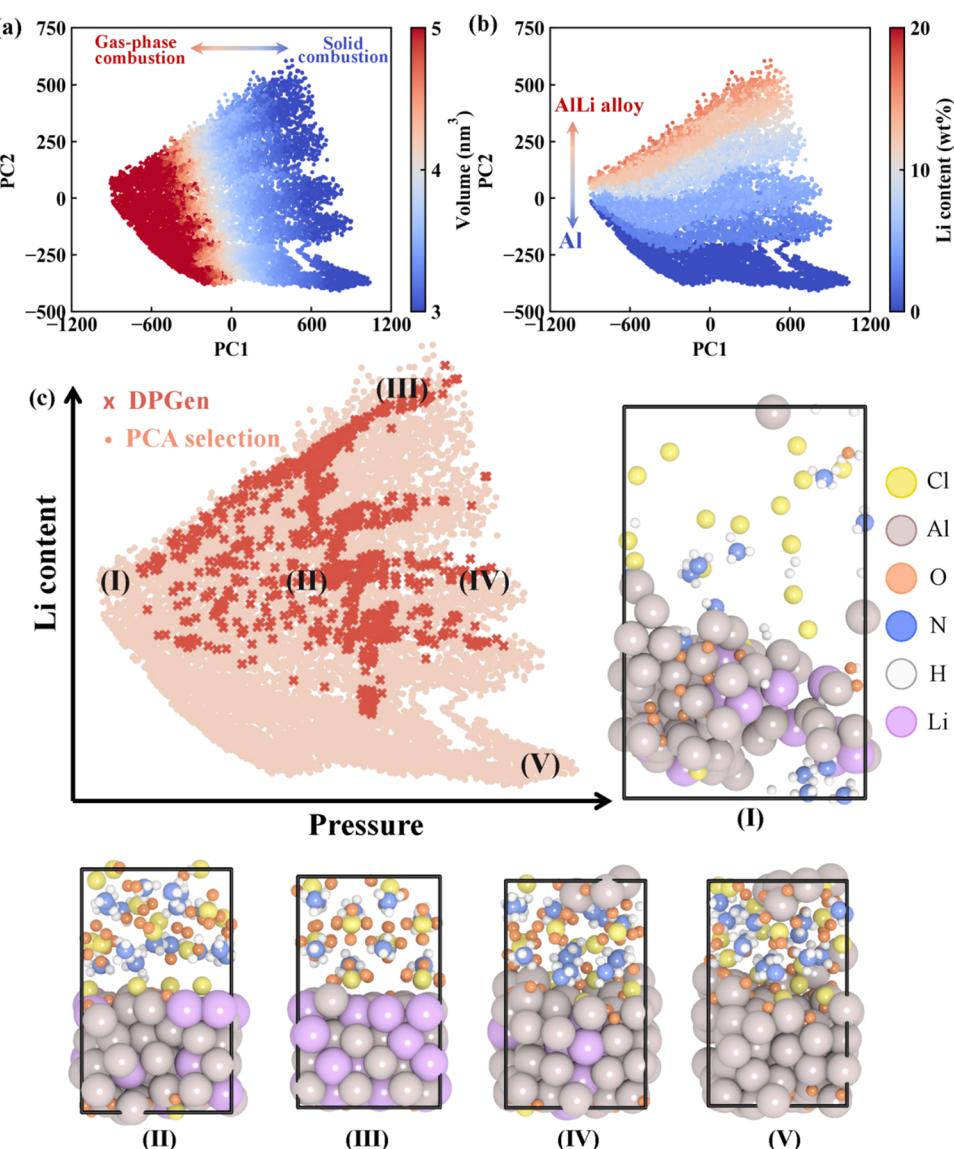
the DP scheme can be found in refs S1,S2. The detailed parameter settings for the training process can be found in the Supporting Information.

Apart from the initial data set, we further apply the active learning algorithm in the DPGen package<sup>20</sup> to ensure the completeness of the data set in configurational space. This technique automatically and iteratively performs three steps: exploration, labeling, and training. Four NNP models are developed using different random seeds. Following this, a short MD simulation is performed using the first NNP model, followed by the prediction of the atomic forces by the remaining three models. Snapshots exhibiting a maximum standard force deviation ( $\delta_F$ ) ranging from 0.3 to 0.5 eV/Å are reassessed and labeled using DFT, and subsequently incorporated into the data set for further training. This sampling strategy involves selecting 255 configurations for the Al-Li system and 1013 configurations for the AlLi-AP system, as outlined in Table 1. All *ab initio* energies and forces are obtained using the CP2K package with the Perdew–Burke–Ernzerhof (PBE) functional form of the generalized gradient approximation to consider the exchange and correlation energies of electrons. The molecular orbitals were expanded using molecularly optimized DZVP-MOLOPT-GTH basis sets. Table 1 summarizes the sampling method and the corresponding size of the data set from each method.

**2.4. Large-Scale Simulation.** The developed NNP model provides a basis for investigating the combustion mechanism of promising alloy-based propellants, with good scalability, retaining the *ab initio* accuracy.<sup>53</sup> Two larger scale configurations (6376 atoms) are established for comparison between the pure metal and its alloyed counterparts, specifically, the Al-AP interface and AlLi-AP interface. The Al<sub>3</sub>Li alloy is employed in the AlLi-AP system, which is the primary phase of the AlLi alloy in experimental investigations. The simulation



**Figure 3.** PCA selection strategy in examining AP decomposition. (a) Landscape of published data set using the PCA technique. (b) Gridding the PCA landscape using a grid size of 10. (c) Data set size when using different grid sizes. A zero value indicates that all data sets were included without selection. (d) Number of decomposed AP molecules predicted by different NNP models, developed by employing various data set sizes. The data set is taken from a previous work.<sup>S4</sup>

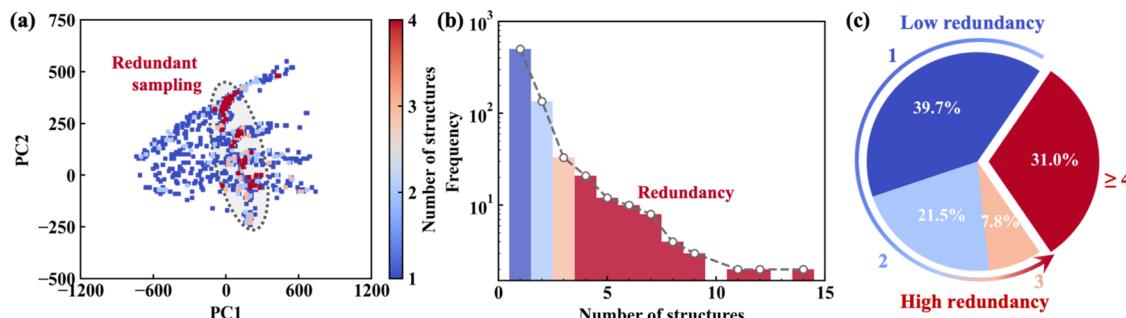


**Figure 4.** PCA projection for sampled configurations, (a) colored by volume, (b) the weight percentage of Li doping in the AlLi alloy, and (c) sampling method. The snapshots in (c) are representative configurations (I–V) on the PCA plane.

box dimensions measure  $2.4 \times 3.7 \times 9.2 \text{ nm}^3$  along three axes, with the metal additive and oxidizer each occupying half of the region in the  $z$ -direction. These configurations represent an amplification of the structure used in the training set. The prepared system was initially equilibrated at 300 K using the NVT ensemble for 10 ps to equilibrate the structure and release residual stresses. In the subsequent process, a combustion process involves fast heating from 300 to 3000 K in 100 ps. The corresponding heat rate is 0.27 K/ps, which is feasible to reveal the thermodynamic behaviors of solid propellant.<sup>54</sup> The time step used in this work corresponds to 0.2 fs, which grants excellent numerical stability over the simulation.<sup>55</sup> A Nose–Hoover thermostat is applied with a dump parameter of 20 fs. All simulations are conducted using LAMMPS,<sup>56</sup> and visualization is performed using OVITO.<sup>57</sup> The data processing and species identification are accomplished using the MDAPY<sup>58</sup> package.

### 3. RESULTS AND DISCUSSIONS

**3.1. PCA Selection.** First of all, the efficiency of our workflow is evaluated using a published data set for pure AP decomposition constructed by Chu et al.<sup>54</sup> The data set comprises 11,826 configurations, with relative energy ranging up to 0.8 eV/atom in Figure 3a. We utilize the SOAP descriptor to construct the attributes of all configurations and analyze their distributions in chemical space using the PCA technique. The first principal component (PC1) and the second principal component (PC2) account for variance proportions of 93 and 6% in the projected features, respectively. This indicates that the two-dimensional projection plane effectively captures the structural characteristic of the data set. The two-dimensional PCA space is further divided by various grid sizes, ranging from 0 to 15. In particular, a grid size of 0 refers to the original data set without selection. An illustration of a grid size of 10 is provided in Figure 3b, and each point refers to a random configuration selected from candidates sharing the same grid. It is notable that over 100 configurations sit in a specific region, displaying high similarity



**Figure 5.** (a) PCA landscape sampled by DPGen with a grid size of 10. (b) Frequency of grids sharing a similar number of configurations. (c) Percentage of the configurations colored by their redundancy. The results in (c) represent the percentage corresponding to the frequency sharing the same level of redundancy.

in configuration. The specific numbers of divided grids and selected configurations are summarized in Figure 3c. Increasing the grid size to 2 sharply reduces the data set size to less than half, specifically 4524 instances. Further increasing the grid size to 15 leads to only 331 configurations in the data set, which accounts for approximately 2.7% of the total configurations available. After selection, six NNP models are trained using varying data set sizes and reproduce the thermal decomposition process of AP across the temperature range of 300–4000 K within 90 ps. Remarkably, all models demonstrate a high level of consistency in the numerical simulations. Notably, even with a small data set of 331 configurations, the model accurately predicts the reaction rate of AP. The streamlined data set, characterized by reduced redundancy and enhanced representability, demonstrates a strong predictive capacity for energy and force of the entire data set (Figure S1). For example, the MAE for energy predictions using a grid size of 10 is only 6.15 meV/atom, marginally higher than the model incorporating the entire data set, which stands at 5.54 meV/atom. The variation corresponds to 10%, which represents the threshold for satisfactory accuracy. Therefore, a grid size of 10 is used to reduce the data set size while maintaining predictive accuracy. These numerical findings offer direct evidence that a set of representative configurations can effectively capture the features of the PES even considering complex reactive interactions, underscoring the importance of utilizing data sets with minimal redundancy in advancing machine learning applications.

The development procedure of the NNP model for AlLi–AP interface follows the workflow outlined in Section 2. Figure 4 presents all sampled configurations in a reduced two-dimensional space. The PC1 and PC2 account for variances of up to 73 and 24% in the projected features, respectively. The two-dimensional projection plane can well capture the inherent structure of the data set, which is consistent with the findings in pure AP decomposition (Figure 3). This section not only presents the PCA plane for the AlLi–AP system, as depicted in Figure 2b, but also includes data from the Al–AP system (Figure 2a) and configurations sampled by the DPGen method. Consequently, there is a slight discrepancy in the PCA distribution compared with Figure 2b. Given the complex dynamics of the interface reaction, it is imperative to discern the structural diversity in the data set. To elucidate the underlying significance of these components, two color modes as volume and Li content are employed. The distinct pattern in Figure 4a establishes that volume significantly influences our data set, illustrating a spectrum from gas-phase combustion to condensed-phase

combustion along the PC1 axis. Li content also plays an important role in distinguishing the sampled data from the PC2 axis on the two-dimensional plane (Figure 4b). In the leftmost area of the PCA plane, the configuration exhibits distinct characteristics regardless of the Li content. This phenomenon is attributed to the failure of the SOAP descriptor for low-pressure conditions that have a larger molecular separation exceeding the current cutoff of the SOAP descriptor as 5 Å. For further analysis, we increased the cutoff values to 6 and 8 Å. It is observed that raising the cutoff leads to a slightly higher explained variance ratio for PC1 and a broader distribution of data points while maintaining the overall shape of the PCA plane, as illustrated in Figure S2 and Table S1. Notably, the 8 Å cutoff value exceeded the commonly used threshold for training NNP models (e.g., 6 Å), indicating consistent interatomic interactions within the gas-phase region. As a result, the computed SOAP information of low-pressure conditions overlaps with other cases.

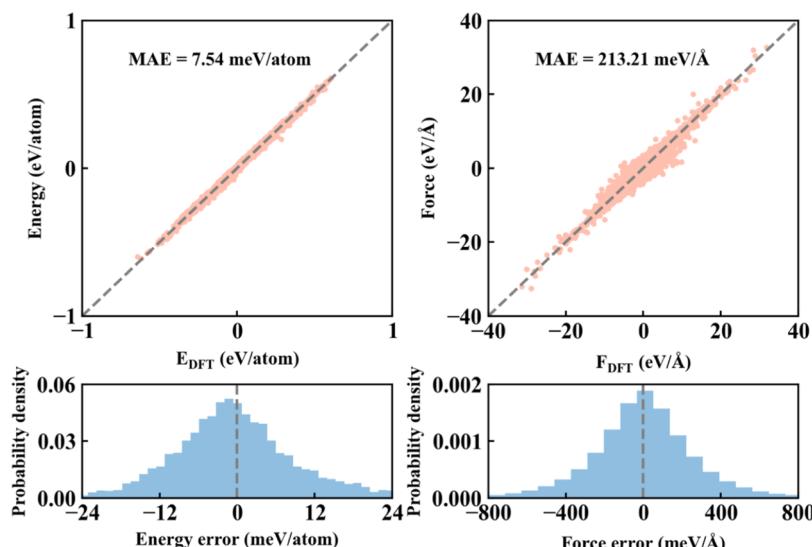
Figure 4c showcases the configurations sampled via the active learning algorithm on the projection plane. It is evident that all configurations sampled by the DPGen method fall within the sampled space of the initial data set. The iterative nature of the DPGen workflow enables a comprehensive exploration of the entire PCA plane, thereby improving the augmentation of underrepresented regions with extra samples. Figure 4c also includes five representative snapshots (I–V) to illustrate the rough differences in configuration and distribution on the PCA plane. The aforementioned findings provide a concrete explanation for the PCA algorithm, enhancing our comprehension of sampling structures.

To characterize the redundancy within the data set, we segmented the PCA landscape into a grid with a size of 10. The PCA selection method effectively reduced the redundancy of the initial data set, as shown in Figure 2, with each grid containing only one configuration. Thus, we illustrate the PCA landscape sampled by DPGen in Figure 5a. During the iterative exploration process, DPGen not only augments the data set but also introduces numerous similar configurations. A total of 1268 configurations sampled by the DPGen method are distributed across 737 grids, with many grids containing more than four configurations, particularly in the central region of Figure 5a. This observation is attributed to the implementation of DPGen and is a consequence of inaccurate predictions for redundant structures in continuous trajectories. Figure 5b illustrates the number of configurations within similar grids and the corresponding frequencies. It is observed that up to 14 configurations share a single grid, indicating high redundancy resulting from active learning. It is generally accepted to

**Table 2.** Data Set Size in Published Works Using a Similar NNP Framework

material	category	<i>T</i>	pressure	reaction	data set size	reference
AlLi	alloy	✓	✓	✗	63,683	Chang et al. <sup>59</sup>
LiCu	alloy	✓	✓	✗	34,885	Lai et al. <sup>60</sup>
AlMgCu	alloy	✓	✓	✗	102,600	Jiang et al. <sup>61</sup>
MgCl <sub>2</sub> –NaCl <sup>a</sup>	interface	✓	✓	✗	27,000	Xu et al. <sup>62</sup>
B–B <sub>2</sub> O <sub>3</sub>	interface	✓	✗	✗	41,970	Chang et al. <sup>53</sup>
AP	molecule	✓	✗	✓	11,826	Chu et al. <sup>54</sup>
ICM-102	molecule	✓	✗	✓	13,966	Chu et al. <sup>31</sup>
CL-20	molecule	✓	✓	✓	81,621	Zhang et al. <sup>63</sup>
Cu–H <sub>2</sub> O <sup>a</sup>	interface	✓	✗	✗	9505	Li et al. <sup>64</sup>
NaCl–SiO <sub>2</sub>	interface	✓	✗	✓	52,000	Zhao et al. <sup>65</sup>
Pt–CO	interface	✓	✗	✓	120,000	Wu et al. <sup>66</sup>
AlLi–AP	interface	✓	✓	✓	11,251	this work

<sup>a</sup>These works only cover a minor temperature range, 280–320 K in Cu–H<sub>2</sub>O and 806–879 K in MgCl<sub>2</sub>–NaCl.



**Figure 6.** Energy and atomic force predictions using the developed NNP model for the training set. Statistical information on prediction errors on energies and forces is shown below.

consider four structures as the critical threshold for high redundancy, and thus, 64 grids are oversampled, approximately 9% of the total. When converting the number of grids into sampled configurations, more than 30% configurations exhibit significant similarity, as depicted in Figure 5c. Despite providing a good starting point, which captures the primary features of the PES, the active learning method still leads to a significant redundancy in the data set. Therefore, it can be reasonably inferred that solely relying on active learning results in generating a significant number of similar structures and imposes a substantial computational burden while supplementing the data set. Based on the above physical findings, the active learning method is recommended to be implemented after PCA selection rather than being independently applied, as illustrated in the workflow in Figure 1. This approach serves as an effective means to alleviate the inherent redundancy within the procedure of active learning procedures.

To highlight the efficacy of the PCA strategy in reducing computational costs, particularly in the context of complex interface reactions, we have assembled a compilation of prior studies utilizing analogous NNP frameworks in Table 2. In terms of the sampling methodology, common techniques involve exploring the configuration space at different temperatures and pressures, potentially including reactions. Con-

sequently, we undertook a thorough examination encompassing these three aspects and the corresponding data set size. The alloy systems encompass sampled configurations of crystal phases and atomic vibrations near equilibrium positions, while investigations of oxidizer molecules involve intricate decomposition reactions. Essentially, the required data set sizes escalate in complexity order from the alloy to the nonreactive interface, oxidant molecule, and complex reactive interface, despite varying sampling conditions. Analysis of interface reactions demands a substantial data set expansion. For example, studying the surface adsorption of CO on the Pt bulk, excluding pressure effects, requires 120,000 data points. This computational scale presents a challenge when analyzing interactions among multiple components in solid propellants across diverse temperature and pressure ranges. An inquiry arises regarding the necessity of such an extensive data set for characterizing configuration space. Through the application of the PCA algorithm, it is discovered that a subset of just 11,251 representative data points can effectively capture the PES of the AlLi–AP interface across various temperature and pressure conditions (Figure 4). In essence, the PCA itself is a practical approach to avoid the financial burden of conducting extensive DFT calculations. It also aids in harnessing the machine

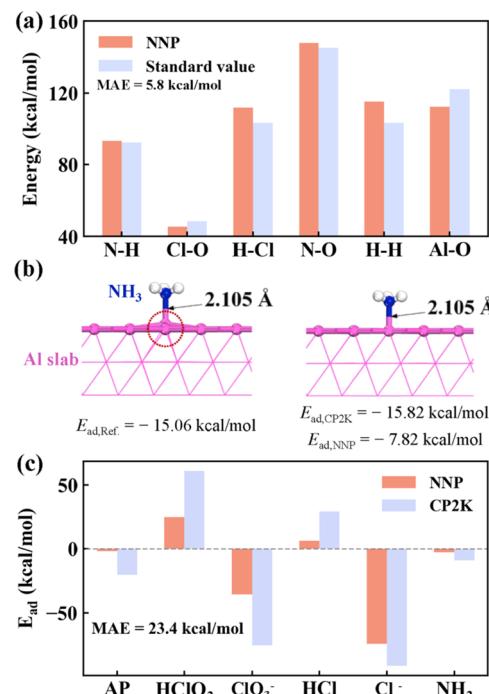
learning potential with a limited number of data points while still upholding accuracy and transferability.

The above investigations employ a published data set as an example and implement the PCA workflow in Figure 1, demonstrating significant potential in minimizing the data requirement while maintaining feasible accuracy. The application of this method in a more intricate AlLi–AP study is then elaborated upon, providing a comprehensive interpretation of PCA distribution. This technique facilitates the identification of representative configurations from a large pool of candidates, resulting in a more streamlined data set with reduced redundancy and data requirements. Such a structural filtering process efficiently captures features on the PES, enabling the detection of rare structures. The approach resembles a sampling approach that measures the similarity of configurations based on their chemical descriptors and can be extended to include additional descriptors such as the Coulomb matrix and atom-centered symmetry function (ACSF) descriptors. Several previous studies<sup>10,31,67</sup> have employed the PCA algorithm to characterize the sampled configurational space, while limited efforts have been made to validate its efficiency and necessity. The active learning method, which learns from mistakes, tends to sample numerous configurations with high similarity from continuous simulation trajectories, thereby increasing the computational burden. Despite a good start, e.g., a representative data set that captures the primary features on the PES, approximately 30% of configurations sampled by DPGen exhibit significant similarity. Therefore, it is advisable for the active learning method to supplement the data set rather than be implemented independently. It serves as a valuable tool to complement the PCA approach, ensuring sampling precision and enhancing model accuracy. Overall, the proposed workflow integrates the classical SOAP descriptor and a practical PCA strategy to minimize redundancy and data requirements, successfully capturing the features of complex PES.

**3.2. Model Validation.** The accuracy of the model predictions is initially validated by assessing the atomic energy and forces, as illustrated in Figure 6. The validation process evaluates the predictive performance of the NNP model on the entire training data set. A distinctive diagonal pattern in the results indicates a high level of accuracy in the developed NNP model. The mean absolute errors (MAE) for atomic energy and force are 7.54 meV/atom and 213.21 meV/Å, respectively. The corresponding error distributions for atomic energy and force fall within  $\pm 20$  meV/atom and  $\pm 600$  meV/Å, respectively. In the research of pure AP decomposition by Chu et al.,<sup>54</sup> the reported MAE for energy and atomic force is 5.55 meV/atom and 184 meV/Å, respectively. Although our model hereby shows slightly higher errors, it encompasses not only the decomposition of AP molecules but also the intricate interface reactions of the AlLi alloy under varying pressure conditions. Moreover, the NNP model demonstrates an energy prediction accuracy of 5–29 meV/atom and a force prediction accuracy of 184–410 meV/Å for single-component substances.<sup>18</sup> Despite being compared to these single substances, our model displays a satisfactory and impressive level of accuracy in predicting intricate interface reactions. Consequently, the coverage of the configurational space in our model is considerably broader compared to that in a previous study. To assess the predictive capability of the NNP model on out-of-sample data, we construct a validation data set for the

Al–AP structure. 200 frames are selected from a heating process spanning from 300 to 4000 K over a duration of 40 ps under the NVT ensemble. A MAE of 15.25 meV/atom is observed for energy prediction and 274.01 meV/Å for force predictions, as illustrated in Figure S3. Compared to the MAE for the training set as 7.54 and 213.21 meV/Å, such satisfactory performance highlights the desirable predictive ability of the developed NNP model on the dynamical reaction within the covered configurational space, as depicted in Figure S4.

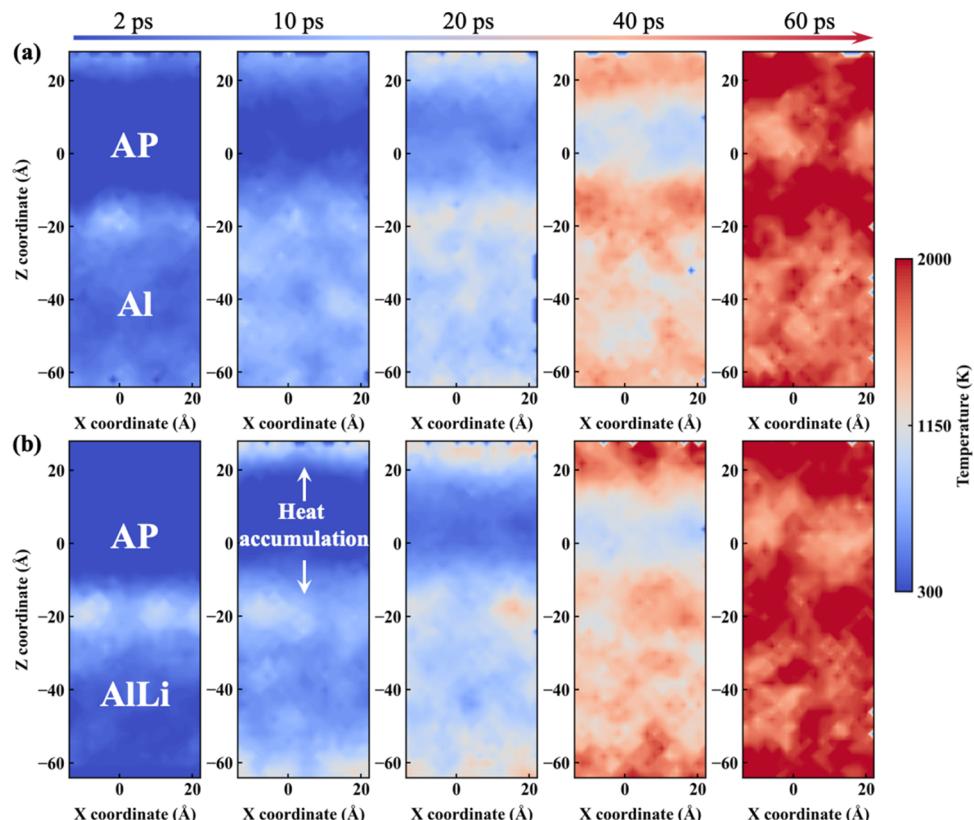
Interface combustion involves bond breakage, formation, and surface reactions. To evaluate the predictive capability of the NNP model in interface interactions, we compare the bond dissociation energies for typical bonds (Figure 7a). Note that



**Figure 7.** (a) Bond dissociation energy for major bonds. (b) Surface adsorption energy of NH<sub>3</sub> on the Al slab and (c) gas molecules on the AlLi slab. The standard values in (a) are extracted from ref 68. The E<sub>ad,ref</sub> in (b) is obtained from ref 71.

the NNP model accounts for bond dissociation in the reaction process at varying pressures and temperatures, without directly incorporating bond dissociation data as done in the ReaxFF parametrization.<sup>6</sup> Unexpectedly, the calculated results of the NNP model align well with the standard values.<sup>68</sup> The maximum difference between these two methods is only 12 kcal/mol in the H–H bond, and the MAE for these six bonds is only 5.8 kcal/mol. The remarkable consistency of the developed NNP model indicates its ability to accurately describe complex interface combustion reactions, which involve bond dissociation and formation. Specifically, a lower bond dissociation energy indicates a higher likelihood of bond breakage. The calculated values suggest that the Cl–O bond in AP is more susceptible to initial bond breakage, reported as a favorable reaction.<sup>69,70</sup>

To evaluate the predictive capability of the developed NNP model beyond the sampled configurational space, we examine the surface adsorptions on the crystal slab that differ from the training data set. The physical adsorption energy (E<sub>ad</sub>) of small

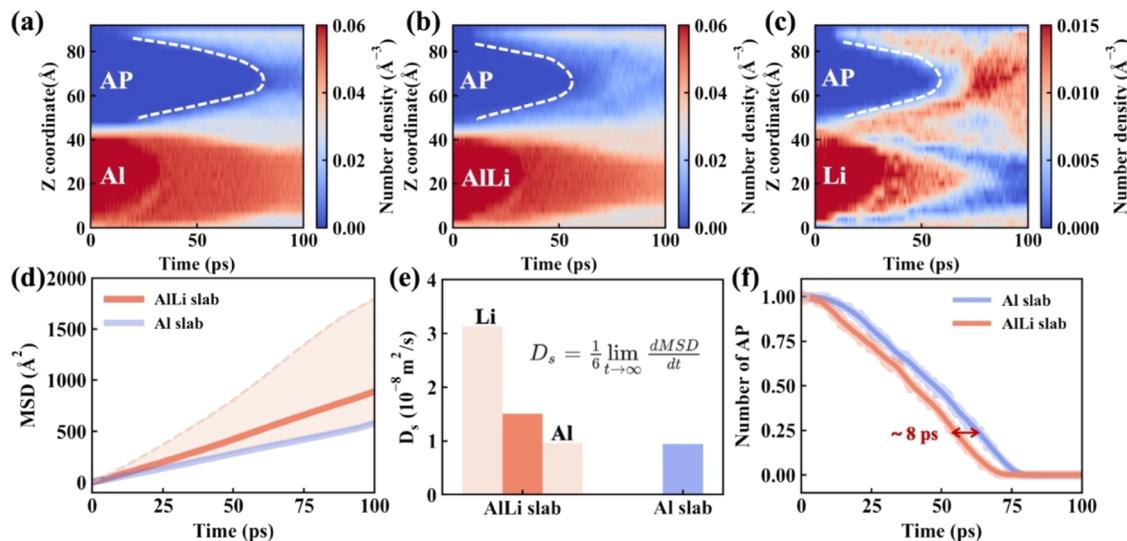


**Figure 8.** Temperature distributions within the (a) Al-AP slab and (b) the Alli-AP slab along reaction propagation in the  $z$ -direction. Two distinct interfaces are presented as a result of periodic boundary conditions, specifically at the top and middle sections of the slabs. The Alli region denotes the  $\text{Al}_3\text{Li}$  crystal, containing 8 wt % Li.

molecules adsorbed on the metal slab is calculated through the following formula:  $E_{\text{ad}} = E_{\text{slab} + \text{molecule}} - E_{\text{slab}} - E_{\text{molecule}}$ . Here,  $E_{\text{slab} + \text{molecule}}$  represents the total energy of the crystal slab with the adsorbed gas molecule, while  $E_{\text{slab}}$  and  $E_{\text{molecule}}$  are the total energies of the metal slab surface and the isolated molecule, respectively. First, we compare the  $E_{\text{ad}}$  for an  $\text{NH}_3$  molecule on an Al slab placed at a distance of 2.105 Å to verify the accuracy of our numerical settings. Previous research<sup>71</sup> reported an  $E_{\text{ad}}$  value of  $-15.06$  kcal/mol for the configuration illustrated in Figure 7b, whereas our calculations using the CP2K approach yielded an  $E_{\text{ad}}$  value of  $-15.82$  kcal/mol. The slight deviation between our findings and published data can be attributed to a minor distortion of the surface Al atom in the reference study. The NNP model yields a markedly lower result, specifically half of the result obtained by CP2K. This difference can be attributed to the fact that this particular configuration was not included in our training set, which emphasized the dynamic interface reactions rather than the ideal interactions between a perfect Al crystal slab and a standard  $\text{NH}_3$  molecule. Besides, the calculated results of key species adsorbed on the Alli slab are illustrated in Figure 7c, employing the CP2K software (PBE-DZVP level) and the customized NNP model. Here, an isolated gas molecule is placed on the  $\text{Al}_3\text{Li}$  alloy slab with a distance of 2 Å, including AP,  $\text{HClO}_3$ ,  $\text{ClO}_3^-$ , HCl,  $\text{Cl}^-$ , and  $\text{NH}_3$ . These species represent the primary intermediate products of AP decomposition, as identified in both numerical simulations and experimental investigations.<sup>54</sup> We include these surface adsorption configurations in the PCA landscape, as depicted in Figure S4. It is observed that the adsorbed configurations deviate significantly from our initial data set,

resulting in an MAE of  $E_{\text{ad}}$  in 23.4 kcal/mol between the NNP model and the CP2K method. Despite the discrepancies between these two methods, the developed NNP model demonstrates a trend similar to the results obtained from CP2K. The above results illustrate that the customized NNP model has its application scope, particularly in solid-phase reaction processes. However, when exceeding the sampled space of training set, the performance of the developed NNP model is limited and conservative. Therefore, it is essential to have a thorough understanding of the data set and its limitations, *e.g.*, with the aid of PCA technique.

**3.3. Interface Combustion.** To investigate the impact of Li doping on the combustion process, the heat and mass transfer *via* the Al-AP and Alli-AP interfaces are monitored. Figure 8 illustrates the temperature variations along the combustion propagation directions, *i.e.*, the  $x$ - $z$  plane. The combustion initiates at the interface between the oxidizer and metal fuels and then spreads toward the interior regions during the fast heating process from 300 to 3000 K in the 100 ps. At 2 ps, the interface temperature in the Al-AP system is observed at  $\sim 600$  K in Figure 8a, which is higher than that of the interior region, indicating a fast interface reaction. In contrast, the interface temperature between the Alli alloy and AP reaches 900 K in Figure 8b, attributed to the significant heat release from surface oxidation reactions. As the reactions proceed to 10 ps, the released heat transfers rapidly within the Al slab, resulting in a uniform temperature distribution due to its higher thermal conductivity. However, Li doping has been found to decrease the thermal conductivity of the alloy, as reported by Chang et al.<sup>59</sup> Specifically, the thermal



**Figure 9.** Temporal evolutions of mass diffusion within (a) Al–AP and (b, c) AlLi–AP slabs. (d) MSD variations for AlLi and Al as well as (e) corresponding self-diffusion coefficient. (f) The number variations for the AP oxidizer in two combustion systems. The boundaries of light pink shadows in panel (d) and the pink bar in panel (e) represent the corresponding results of Li and Al atoms within the AlLi–AP system. The AlLi region denotes the  $\text{Al}_3\text{Li}$  crystal, containing 8 wt % of Li.

conductivity of Al is  $241.5 \text{ W}/(\text{m}\cdot\text{K})$  whereas that of  $\text{Al}_3\text{Li}$  is  $101.9 \text{ W}/(\text{m}\cdot\text{K})$ , resulting in significantly slower heat transfer within the Al–Li alloy. The heat generated from the reaction accumulates at the AlLi–AP interface, thereby reducing the ignition delay of the system. As the interface combustion further progresses, the atomic temperature distributions within the two systems behave similarly. The temperature gradient in the AP region diminishes gradually, leading to a uniform distribution of the temperature throughout the system at 60 ps. The thermal analyses presented above offer atomistic evidence of the impact of Li doping on the combustion interface between the metal and the oxidizer. This evidence includes a substantial release of reaction heat and increased heat accumulation, attributed to the superior active ability and lower thermal conductivity. These factors play a crucial role in enhancing the initial combustion of solid propellants.

The propagation of the combustion front is tightly interconnected with mass transfer in the Al–AP and AlLi–AP systems. The diffusion of the metal atom along the z-axis, which corresponds to the reaction front, is examined. Figure 9a displays the temporal changes in the number density of Al within the Al–AP system, and Figure 9b,c illustrates the evolution of the AlLi and Li in the AlLi–AP system. Dashed lines in the figures highlight the diffusion front of metal atoms, reflecting the burning rate within solid propellant. Observations in the Al–AP system reveal that Al atoms penetrate the central region of the AP slab in ~80 ps, whereas the AlLi atoms in the AlLi–AP system achieve this penetration in ~55 ps. The incorporation of Li doping enhances the combustion propagation by 30% due to its superior diffusivity. To quantify the diffusion of metal additives, the mean square displacements (MSD) are summarized in Figure 9d. The MSD values are calculated at every 1 ps, representing the deviation of the atom position with respect to its reference position over the simulation time. The results suggest that the atom movement within AlLi alloy is significantly faster than that within the Al slab. Figure 9e presents the self-diffusion coefficient ( $D_s$ ) of metal atoms, calculated by the corresponding slope of MSD curves. To differentiate the contributions of Al and Li atoms

within the AlLi–AP system, we also determined the individual diffusion rates are also determined. It is confirmed that the diffusion coefficient of Li is 3 times that of Al, owing to its small atomic size. A similar trend is observed for the double-size configuration, with the diffusion coefficient being 2.8 times larger, as illustrated in Figure S5. This high level of consistency indicates that the current numerical configuration size is well-suited for examining mass diffusion within the system. Thus, as the amount of Li doping increases, the overall mass diffusion rate of the alloy also rises, leading to a higher burning rate of the metal additive. Furthermore, the diffusion rates of Al atoms in both systems are consistent, indicating that the faster ignition in the AlLi–AP system is driven by Li diffusion.

The enhanced mass diffusion of the metal additive accelerates the decomposition of the oxidizer component. Figure 9f illustrates the evolution of the number of AP molecules in the Al–AP and AlLi–AP systems. It is evident that the overall decomposition process is accelerated by ~8 ps when replacing an Al slab with an AlLi alloy. This acceleration corresponds to a 10% enhancement in the AlLi–AP interface combustion. The experimental study noted a swift initiation of exothermic decomposition at 683 K in the Al–AP system and at 633 K in the AlLi–AP system.<sup>38</sup> The entire thermal decomposition process was enhanced by 7.3%, which is consistent with the increased burning rate observed in our numerical investigation. Furthermore, Figure S6 illustrates the predominant gas-phase species during the decomposition of AP at two different interfaces. The results indicate that Li doping does not introduce new species but slightly reduces their quantities. The above NNP-based numerical findings provide atomistic evidence of Li doping on the heat and mass transfer, as well as the combustion process within solid propellant.

#### 4. CONCLUSIONS

Efficiently sampling the configurational space with a limited data set poses a significant challenge in the development of machine learning models that aim for high efficiency and accuracy. This study employs the SOAP descriptor and PCA

strategy to characterize the features of sampled structures generated by MD simulations, initially employing a low-precision force field. The PCA algorithm enables selecting the representative configurations to represent the intricate PES. This approach demonstrates excellent performance when tested against published works, offering an effective way to sample the configurational space with low redundancy and limited data. Subsequently, the active learning method is employed to supplement the data set, thereby refining and fine-tuning the model's ability to describe the PES. With this proposed workflow, the study develops an NNP model using only 11,251 configurations, particularly suitable for modeling complex interface reactions across a wide range of temperatures and pressures. The developed model exhibits good predictive capabilities in predicting energy, atomic force, and chemical bond energy. A series of large-scale MD simulations are performed, conducted on the interface combustion within solid propellants, focusing on the Al–AP and AlLi–AP interfaces. Investigations of heat transfer within the slab reveal that Li doping significantly influences the initial combustion stage, enhancing reactivity and reducing thermal conductivity. Mass transfer analysis highlights the considerably higher diffusion coefficient of Li compared to Al, with the former being 3 times greater. Consequently, the overall combustion process is accelerated by ~10%. The above findings offer compelling evidence at the atomistic level for the improved combustion performance of the AlLi alloy. These pioneering efforts represent a breakthrough in the utilization of alloy-formula propellants, establishing a foundational framework for virtual screening and the rational design of innovative propellant formulations and microstructures.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

Data will be made available on request.

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.4c00587>.

Detailed numerical settings for methodology; MAE for energy and force predictions employing NNP models across different grid sizes; PCA plane for sampled configurations using different descriptor cutoff; energy and atomic force predictions using the developed NNP model on the validation data set; PCA plane for the training data set, validation data set, and surface adsorption configurations; self-diffusion coefficient for AlLi–AP systems with different size; major gas-phase species during AP decomposition in two systems; explained variance ratio and distributions of the PCA plane using different descriptor cutoff. ([DOCX](#))

Scripts for PCA selection and NNP training. ([ZIP](#))

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Qingzhao Chu** – State Key Laboratory of Explosion Science and Technology, Beijing Institute of Technology, Beijing 100081, China; [ORCID iD](https://orcid.org/0000-0002-2090-1864); Email: [chuqz@bit.edu.cn](mailto:chuqz@bit.edu.cn)

**Dongping Chen** – State Key Laboratory of Explosion Science and Technology, Beijing Institute of Technology, Beijing 100081, China; [ORCID iD](https://orcid.org/0000-0002-4232-3356); Email: [dc516@bit.edu.cn](mailto:dc516@bit.edu.cn)

## Authors

**Xiaoya Chang** – State Key Laboratory of Explosion Science and Technology, Beijing Institute of Technology, Beijing 100081, China; [ORCID iD](https://orcid.org/0009-0009-1941-4602)

**Di Zhang** – State Key Laboratory of Explosion Science and Technology, Beijing Institute of Technology, Beijing 100081, China

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.4c00587>

## Author Contributions

X.C.: Investigation, data curation, formal analysis, methodology, visualization, and writing – original draft. D.Z.: Methodology and validation. Q.C.: Conceptualization, methodology, validation, and writing – review and editing. D.C.: Conceptualization, funding acquisition, supervision, and writing – review and editing. All authors discussed the results, and contributed to the manuscript writing and approved the final version.

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work is supported by the State Key Laboratory of Explosion Science and Technology (Grant ZDKT21-01), the National Natural Science Foundation of China (Grant S2106130), and the Science and Technology Innovation Program of Beijing Institute of Technology (2022CX01028).

## ■ REFERENCES

- (1) Jain, A.; Shin, Y.; Persson, K. A. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.* **2016**, *1* (1), 15004.
- (2) Miró, P.; Audiffred, M.; Heine, T. An atlas of two-dimensional materials. *Chem. Soc. Rev.* **2014**, *43* (18), 6537–6554.
- (3) Hospital, A.; Goni, J. R.; Orozco, M.; Gelpí, J. L. Molecular dynamics simulations: Advances and applications. *Adv. Appl. Bioinform. Chem.* **2015**, *8*, 37–47.
- (4) Carloni, P.; Rothlisberger, U.; Parrinello, M. The role and perspective of *ab initio* molecular dynamics in the study of biological systems. *Acc. Chem. Res.* **2002**, *35* (6), 455–464.
- (5) Ramachandran, K.; Deepa, G.; Namboori, K. *Computational Chemistry and Molecular Modeling: Principles and Applications*; Springer Science & Business Media, 2008.
- (6) Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; et al. The reaxff reactive force-field: development, applications and future directions. *npj Comput. Mater.* **2016**, *2* (1), 15011.
- (7) Friederich, P.; Hase, F.; Proppe, J.; Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. *Nat. Mater.* **2021**, *20* (6), 750–761.
- (8) Pinheiro, M., Jr.; Ge, F.; Ferre, N.; Dral, P. O.; Barbatti, M. Choosing the right molecular machine learning potential. *Chem. Sci.* **2021**, *12* (43), 14396–14413.
- (9) Bayerl, D.; Andolina, C. M.; Dwaraknath, S.; Saidi, W. A. Convergence acceleration in machine learning potentials for atomistic simulations. *Digital Discovery* **2022**, *1* (1), 61–69.
- (10) Noordhoek, K.; Bartel, C. J. Accelerating the prediction of inorganic surfaces with machine learning interatomic potentials. *Nanoscale* **2024**, *16* (13), 6365–6382.
- (11) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559* (7715), 547–555.

- (12) Wieder, M.; Fass, J.; Chodera, J. D. Fitting quantum machine learning potentials to experimental free energy data: Predicting tautomer ratios in solution. *Chem. Sci.* **2021**, *12* (34), 11364–11381.
- (13) Deringer, V. L.; Caro, M. A.; Csanyi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **2019**, *31* (46), No. 1902765.
- (14) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schutt, K. T.; Tkatchenko, A.; Muller, K. R. Machine learning force fields. *Chem. Rev.* **2021**, *121* (16), 10142–10186.
- (15) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145* (17), No. 170901.
- (16) Zhou, Y.; Ouyang, Y.; Zhang, Y.; Li, Q.; Wang, J. Machine learning assisted simulations of electrochemical interfaces: Recent progress and challenges. *J. Phys. Chem. Lett.* **2023**, *14* (9), 2308–2316.
- (17) Fonseca, G.; Poltavsky, I.; Vassilev-Galindo, V.; Tkatchenko, A. Improving molecular force fields across configurational space by combining supervised and unsupervised machine learning. *J. Chem. Phys.* **2021**, *154* (12), No. 124102.
- (18) Chang, X.; Wen, M.; Zhang, D.; Wang, Y.; Chu, Q.; Zhu, T.; Chen, D. Recent progress toward molecular modeling of energetic materials by using machine learning potential. *Chin. J. Explos. Propellants* **2023**, *45*, 361–377.
- (19) Behler, J. Neural network potential-energy surfaces in chemistry: A tool for large-scale simulations. *Phys. Chem. Chem. Phys.* **2011**, *13* (40), 17930–17955.
- (20) Zhang, Y.; Wang, H.; Chen, W.; Zeng, J.; Zhang, L.; Wang, H.; Weinan, E. DP-Gen: A concurrent learning platform for the generation of reliable deep learning based potential energy models. *Comput. Phys. Commun.* **2020**, *253*, No. 107206.
- (21) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less is more: sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148* (24), No. 241733.
- (22) Xu, L.; Shao, W.; Jin, H.; Wang, Q. Data efficient and stability indicated sampling for developing reactive machine learning potential to achieve ultralong simulation in lithium-metal batteries. *J. Phys. Chem. C* **2023**, *127* (50), 24106–24117.
- (23) Magar, R.; Farimani, A. B. Learning from mistakes: Sampling strategies to efficiently train machine learning models for material property prediction. *Comput. Mater. Sci.* **2023**, *224*, No. 112167.
- (24) Zhang, H.; Juraskova, V.; Duarte, F. Modeling chemical processes in explicit solvents with machine learning potentials. *ChemRxiv* **2024**, DOI: 10.26434/chemrxiv-2023-ktsqc-v2.
- (25) Finkbeiner, J.; Tovey, S.; Holm, C. Generating minimal training sets for machine learned potentials. *Phys. Rev. Lett.* **2024**, *132* (16), No. 167301.
- (26) Mukherjee, K.; Colón, Y. J. Machine learning and descriptor selection for the computational discovery of metal-organic frameworks. *Mol. Simul.* **2021**, *47* (10–11), 857–877.
- (27) Imbalzano, G.; Anelli, A.; Giofre, D.; Klees, S.; Behler, J.; Ceriotti, M. Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials. *J. Chem. Phys.* **2018**, *148* (24), No. 241730.
- (28) Xu, M.; Zhu, T.; Zhang, J. Z. H. Automated construction of neural network potential energy surface: The enhanced self-organizing incremental neural network deep potential method. *J. Chem. Inf. Model.* **2021**, *61* (11), 5425–5437.
- (29) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemonette, L. Enhanced Sampling Methods for Molecular Dynamics Simulations. *Living J. Comp. Mol. Sci.* **2022**, *4*, 1583.
- (30) Gao, Y.; Lu, Y.; Zhu, X. Mateverse, the future materials science computation platform based on metaverse. *J. Phys. Chem. Lett.* **2023**, *14* (1), 148–157.
- (31) Chu, Q.; Luo, K. H.; Chen, D. Exploring complex reaction networks using neural network-based molecular dynamics simulation. *J. Phys. Chem. Lett.* **2022**, *13* (18), 4052–4057.
- (32) Wan, K.; He, J.; Shi, X. Construction of high accuracy machine learning interatomic potential for surface/interface of nanomaterials-a review. *Adv. Mater.* **2024**, *36*, 2305758.
- (33) Chang, X.; Chu, Q.; Chen, D. Shock-induced anisotropic metal combustion. *J. Phys. Chem. C* **2020**, *124* (24), 13206–13214.
- (34) Belal, H.; Han, C. W.; Gunduz, I. E.; Ortalan, V.; Son, S. F. Ignition and combustion behavior of mechanically activated Al–Mg particles in composite solid propellants. *Combust. Flame* **2018**, *194*, 410–418.
- (35) Zhu, Y.; Le, W.; Zhao, W.; Ma, X.; Liu, D.; Li, J.; Jiao, Q. Promising fuels for energetics: Spherical Al-Li powders with high reactivity via incorporation of Li. *Fuel* **2022**, *323*, No. 124393.
- (36) Schiemann, M.; Bergthorson, J.; Fischer, P.; Scherer, V.; Tarota, D.; Schmid, G. A review on lithium combustion. *Appl. Energy* **2016**, *162*, 948–965.
- (37) Sundaram, D. S.; Puri, P.; Yang, V. A general theory of ignition and combustion of nano- and micron-sized aluminum particles. *Combust. Flame* **2016**, *169*, 94–109.
- (38) Terry, B. C.; Sippel, T. R.; Pfeil, M. A.; Gunduz, I. E.; Son, S. F. Removing hydrochloric acid exhaust products from high performance solid rocket propellant using aluminum-lithium alloy. *J. Hazard. Mater.* **2016**, *317*, 259–266.
- (39) Terry, B. C.; Gunduz, I. E.; Pfeil, M. A.; Sippel, T. R.; Son, S. F. A mechanism for shattering microexplosions and dispersive boiling phenomena in aluminum–lithium alloy based solid propellant. *Proc. Combust. Inst.* **2017**, *36* (2), 2309–2316.
- (40) Chu, Q.; Fu, X.; Zhen, X.; Liu, J.; Chen, D. Molecular simulations of HTPB/Al/AP/RDX propellants combustion. *Chin. J. Explos. Propellants* **2024**, *47* (3), 254–261.
- (41) Wang, H.; Rehwoldt, M. C.; Wang, X.; Yang, Y.; Zachariah, M. R. On the promotion of high temperature ap decomposition with silica mesoparticles. *Combust. Flame* **2019**, *200*, 296–302.
- (42) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87* (18), No. 184115.
- (43) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Canova, F. F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. Dscribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **2020**, *247*, No. 106949.
- (44) Manna, S.; Loeffler, T. D.; Batra, R.; Banik, S.; Chan, H.; Varughese, B.; Sasikumar, K.; Sternberg, M.; Peterka, T.; Cherukara, M. J.; et al. Learning in continuous action space for developing high dimensional potential energy models. *Nat. Commun.* **2022**, *13* (1), No. 368.
- (45) Jäger, M. O. J.; Morooka, E. V.; Canova, F. F.; Himanen, L.; Foster, A. S. Machine learning hydrogen adsorption on nanoclusters through structural descriptors. *npj Comput. Mater.* **2018**, *4* (1), 37.
- (46) Gardin, A.; Perego, C.; Doni, G.; Pavan, G. M. Classifying soft self-assembled materials via unsupervised machine learning of defects. *Commun. Chem.* **2022**, *5* (1), 82.
- (47) Zhang, L.; Han, J.; Wang, H.; Car, R.; Weinan, E. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **2018**, *120* (14), No. 143001.
- (48) Calegari Andrade, M. F.; Ko, H. Y.; Zhang, L.; Car, R.; Selloni, A. Free energy of proton transfer at the water-TiO<sub>2</sub> interface from ab initio deep potential molecular dynamics. *Chem. Sci.* **2020**, *11* (9), 2335–2341.
- (49) Liu, S.; Dupuis, R.; Fan, D.; Benzaria, S.; Bonneau, M.; Bhatt, P.; Eddaudia, M.; Maurin, G. Machine learning potential for modelling H<sub>2</sub> adsorption/diffusion in MOFs with open metal sites. *Chem. Sci.* **2024**, *15* (14), 5294–5302.
- (50) Lu, D.; Wang, H.; Chen, M.; Lin, L.; Car, R.; Weinan, E.; Jia, W.; Zhang, L. 86 pflops deep potential molecular dynamics simulation of 100 million atoms with *ab initio* accuracy. *Comput. Phys. Commun.* **2021**, *259*, No. 107624.
- (51) Wang, H.; Zhang, L.; Han, J.; Weinan, E. Deepmd-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178–184.
- (52) Zeng, J.; Zhang, D.; Lu, D.; Mo, P.; Li, Z.; Chen, Y.; Rynik, M.; Huang, L.; Li, Z.; Shi, S.; et al. DeepMD-kit v2: a software package for deep potential models. *J. Chem. Phys.* **2023**, *159* (5), No. 054801.

- (53) Chang, X.; Chu, Q.; Chen, D. Monitoring the melting behavior of boron nanoparticles using a neural network potential. *Phys. Chem. Chem. Phys.* **2023**, *25* (18), 12841–12853.
- (54) Chu, Q.; Wen, M.; Fu, X.; Eslami, A.; Chen, D. Reaction network of ammonium perchlorate (AP) decomposition: The missing piece from atomic simulations. *J. Phys. Chem. C* **2023**, *127* (27), 12976–12982.
- (55) Pang, K.; Wen, M.; Chang, X.; Xu, Y.; Chu, Q.; Chen, D. The thermal decomposition mechanism of RDX/AP composites: *ab initio* neural network md simulations. *Phys. Chem. Chem. Phys.* **2024**, *26* (15), 11545–11557.
- (56) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; et al. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, No. 108171.
- (57) Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO—the open visualization tool. *Model. Simul. Mater. Sci. Eng.* **2010**, *18* (1), No. 015012.
- (58) Wu, Y.-C.; Shao, J.-L. Mdapy: A flexible and efficient analysis software for molecular dynamics simulations. *Comput. Phys. Commun.* **2023**, *290*, No. 108764.
- (59) Chang, X.; Wu, Y.; Chu, Q.; Zhang, G.; Chen, D. *Ab initio* driven exploration on the thermal properties of Al-Li alloy. *ACS Appl. Mater. Interfaces* **2024**, *16*, 14954–14964.
- (60) Lai, G.; Jiao, J.; Fang, C.; Zhang, R.; Xu, X.; Sheng, L.; Jiang, Y.; Ouyang, C.; Zheng, J. A deep neural network interface potential for Li-Cu systems. *Adv. Mater. Interfaces* **2022**, *9* (27), No. 2201346.
- (61) Jiang, W.; Zhang, Y.; Zhang, L.; Wang, H. Accurate deep potential model for the Al–Cu–Mg alloy in the full concentration space. *Chin. Phys. B* **2021**, *30* (5), No. 050706.
- (62) Xu, T.; Li, X.; Wang, Y.; Tang, Z. Development of deep potentials of molten MgCl<sub>2</sub>-NaCl and MgCl<sub>2</sub>-KCl salts driven by machine learning. *ACS Appl. Mater. Interfaces* **2023**, *15*, 14184–14195.
- (63) Zhang, J.; Guo, W.; Yao, Y. Deep potential molecular dynamics study of chapman-jouguet detonation events of energetic materials. *J. Phys. Chem. Lett.* **2023**, *14* (32), 7141–7148.
- (64) Li, Z.; Tan, X.; Fu, Z.; Liu, L.; Yang, J. Y. Thermal transport across copper-water interfaces according to deep potential molecular dynamics. *Phys. Chem. Chem. Phys.* **2023**, *25* (9), 6746–6756.
- (65) Zhao, C. Y.; Tao, Y. B.; He, Y. Microstructure and thermophysical property prediction for chloride composite phase change materials: a deep potential molecular dynamics study. *J. Phys. Chem. C* **2023**, *127* (14), 6852–6860.
- (66) Wu, J.; Chen, D.; Chen, J.; Wang, H. Structural and composition evolution of palladium catalyst for CO oxidation under steady-state reaction conditions. *J. Phys. Chem. C* **2023**, *127* (13), 6262–6270.
- (67) Shkurti, A.; Styliari, I. D.; Balasubramanian, V.; Bethune, I.; Pedebos, C.; Jha, S.; Laughton, C. A. CoCo-MD: A simple and effective method for the enhanced sampling of conformational space. *J. Chem. Theory Comput.* **2019**, *15* (4), 2587–2596.
- (68) Szwarc, M. Bond dissociation energies. *Q. Rev. Chem. Soc.* **1951**, *5* (1), 22–43.
- (69) Zhu, W.; Wei, T.; Zhu, W.; Xiao, H. Comparative DFT study of crystalline ammonium perchlorate and ammonium dinitramide. *J. Phys. Chem. A* **2008**, *112*, 4688–4693.
- (70) Patel, J.; Panchal, H.; Chowdhury, A.; Kumbhakarna, N. Initiation step in the condensed phase decomposition process of ammonium perchlorate. *ChemistrySelect* **2023**, *8* (46), No. e202303409.
- (71) Yang, Q.; Xu, R.; Nie, H.; Yan, Q.; Liu, J.; Chen, J.; Sun, Y. The effects of Al (111) and Al<sub>2</sub>O<sub>3</sub>/Al (111) surfaces on the decomposition of ammonium perchlorate: a DFT-D study. *Mol. Catal.* **2024**, *553*, No. 113712.