







Cite this: DOI: 10.1039/d6ta01407k

Material characterization of aluminosilicate hydrate geopolymers using deep learning assisted tailor-made potential

Yu Li,  Jia-ao Hou,  Yuqi Feng, Danyang Zhao, Cheuk Lun Chow * and Denvid Lau *

Aluminosilicate hydrate geopolymers, such as sodium aluminosilicate hydrate and potassium aluminosilicate hydrate, are low-carbon cementitious materials commonly synthesized from industrial solid wastes such as fly ash and slag, offering a sustainable cementitious material with superior mechanical and environmental performance. However, their atomic-scale mechanisms remain elusive due to limitations in experimental and conventional simulation methods. In this study, a machine learning potential model constructed within a deep potential generator framework is developed for sodium aluminosilicate hydrate, and trained on density functional theory datasets spanning from 300 K to 1000 K. The model accurately reproduces density functional theory-calculated energies and forces with errors of 0.005 eV per atom and 0.078 eV Å⁻¹. Furthermore, transfer learning is employed to adapt the sodium aluminosilicate hydrate model to potassium aluminosilicate hydrate using a small amount of additional density functional theory data, yielding comparable accuracy and faster convergence, with errors of 0.003 eV per atom and 0.092 eV Å⁻¹ for energies and forces. To the best of our knowledge, this work presents the first machine-learning interatomic potentials specifically developed for both N-A-S-H and K-A-S-H geopolymer gels. Structural characterization, elastic properties, and dynamic behaviors predicted by the models are benchmarked against density functional theory, classical forcefields, and experimental measurements, demonstrating robustness and transferability of the approach. The findings demonstrate that the method is highly capable of reliably capturing complex aluminosilicate systems, which provide a new atomic-level understanding of their structural and mechanical behavior, thereby establishing a robust basis for guiding the targeted design of durable, high-performance, and sustainable geopolymer materials.

Received 14th February 2026

Accepted 29th May 2026

DOI: 10.1039/d6ta01407k

rsc.li/materials-a

1. Introduction

Geopolymers have gained increasing recognition as an important category of advanced inorganic binder materials and are commonly synthesized *via* alkaline activation processes.^{1,2} This process involves the chemical reaction between aluminosilicate-rich precursor sources, including materials such as metakaolin, slag, fly ash, and various aluminosilicate-rich precursors, with hydroxide or silicate-based alkaline activators.^{3,4} The resulting geopolymers feature a distinctive amorphous three-dimensional network architecture, generated through the polycondensation of silicate and aluminate tetrahedral units linked *via* oxygen-bridge bonding connections.⁵ This unique structural configuration endows geopolymers with exceptional mechanical performance and enhanced durability, with certain properties surpassing those of conventional Portland cement.⁶ Among the reported diverse geopolymer systems, sodium

aluminosilicate hydrate (N-A-S-H) represents a commonly encountered gel-forming component, typically expressed by the formula $n[\text{Na}_2\text{O} \cdot x\text{SiO}_2 \cdot \text{Al}_2\text{O}_3 \cdot y\text{H}_2\text{O}]$, where n denotes the polymerization degree, and x and y vary with the phase-specific stoichiometric composition, as shown in Fig. 1. In addition to N-A-S-H, potassium aluminosilicate hydrate (K-A-S-H) and analogous systems have also become focal points of extensive research endeavors, as shown in Fig. 1. The widespread application of geopolymers confers multiple benefits. In terms of resource utilization, the synthesis of geopolymers offers significant advantages in resource utilization, as it enables the valorization of industrial residues enriched in aluminum and silicon elements, as well as calcined clay materials.⁷ This not only addresses the challenge of solid waste management but also offers an efficient route for immobilizing and stabilizing hazardous heavy-metal species within the three-dimensional geopolymer network matrix.⁸ From an environmental standpoint, geopolymer production exhibits pronounced sustainability advantages, with energy consumption reduced by approximately 70% and carbon dioxide emissions lowered by

Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, China. E-mail: denvid.lau@cityu.edu.hk; cheuchow@cityu.edu.hk



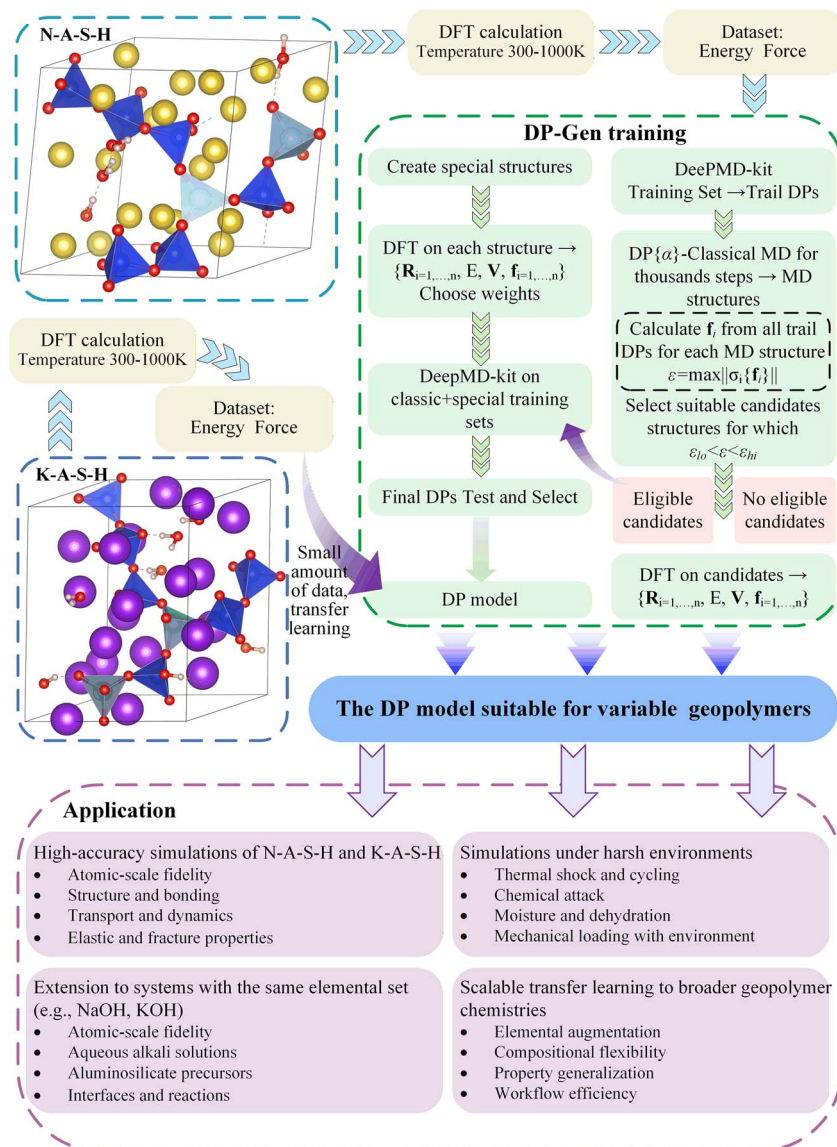


Fig. 1 Workflow for developing a DPMD model tailored to variable geopolymer systems, integrating dataset preparation, model training, and application.

up to 80% relative to conventional Portland cement manufacturing routes.^{9–11}

Although geopolymers show tremendous promise in terms of material performance and environmental applications, their formation mechanisms and microstructural evolution remain poorly understood. Conventional characterization approaches, including X-ray diffraction and nuclear magnetic resonance spectroscopy, are largely limited to inferring atomic-scale structural features in an indirect manner, while dynamic processes such as ion migration and hydrogen bond formation are still beyond direct observation. High-resolution characterization methods are not only expensive but also difficult to implement widely. Against this backdrop, classical molecular dynamics (MD) simulations offer an effective computational framework for investigating thermodynamic properties, structural units, and dynamic behaviors at the atomic level.^{12–18} These simulations contribute to unveiling the formation

mechanisms of geopolymers and guiding performance optimization. Prior studies employing non-reactive Buckingham or Morse forcefields and reactive forcefield (ReaxFF) have constructed three-dimensional sodium aluminosilicate (N-A-A) glass or network models under constant-energy, constant-pressure and constant-temperature ensembles.^{19–21} These investigations reveal, for instance, that increasing water content from 2.6% to 4.7% increases the sodium diffusion coefficient, from $7.82 \times 10^{-13} \text{ m}^2 \text{ s}^{-1}$ to $10.12 \times 10^{-13} \text{ m}^2 \text{ s}^{-1}$, that models with silicon-to-aluminum ratios of two and three exhibit the highest tensile strength, and that noncompact models containing nanopores and water display strengths 45% lower than that of their compact analogues.²² Additional ReaxFF simulations have demonstrated that high temperatures accelerate network polymerization, that silicon-oxygen and aluminum-oxygen bond lengths first increase then decrease, and that magnesium binds more strongly to the aluminosilicate



framework than sodium while cesium shows the highest mobility.^{22,23} Nevertheless, empirical forcefields confront an inherent trade-off between computational efficiency and chemical fidelity. Furthermore, *ab initio* molecular dynamics (AIMD) simulations, which generate interatomic forces *via* density functional theory (DFT), can deliver unparalleled precision in describing bonding and reaction events.^{24,25} However, their practical applicability is typically restricted to systems containing only a few hundred atoms and simulation times on the order of tens of picoseconds, which severely limits their use for large-scale geopolymer dynamics.

Machine learning potential (MLP) has recently been recognized as an effective approach for achieving a balance between accuracy and computational efficiency in MD simulations. Trained on first-principles datasets, these models can achieve quantum-level accuracy without incurring prohibitive computational cost.^{26,27} Deep potential (DP), a multilayer perceptron framework, excels at capturing complex nonlinear relationships between atomic coordinates and system energy.²⁸ Open-source packages such as deep potential molecular dynamics (DPMD) based on passive learning and Deep Potential Generator (DP-GEN) employing concurrent learning have been developed to facilitate model construction.²⁹ The DP-GEN concurrent-learning strategy iteratively refines the training set by screening MD trajectories for force deviations outside predefined bounds and supplementing data *via* single-point first-principles calculations.²⁸ Although this methodology has propelled advances in calcium silicate hydrate research, its application to geopolymer systems remains unexplored, underscoring the need for dedicated machine learning potential studies. More importantly, the unresolved challenge is not only whether existing MLP frameworks can be transferred to geopolymer systems, but whether they can be used to construct system-specific interatomic potentials for the particular amorphous aluminosilicate gels that researchers actually aim to understand and design. In conventional forcefield approaches, once a parameter set is fixed, extension to related but compositionally different systems often requires reparameterization, empirical adjustment, or the use of separate and potentially inconsistent descriptions. A MLP framework offers the opportunity to overcome this limitation by providing a more flexible yet still first-principles-grounded route for building coherent potentials tailored to the systems of interest.

This study aims to formulate a DPMD model with applicability to sodium-based and potassium-based aluminosilicate geopolymers. Starting from established sodium aluminosilicate glass structures, aluminum atoms are randomly substituted by silicon and charge neutrality is maintained through the incorporation of calcium cations. Subsequent hydration yields realistic N-A-S-H frameworks, which are geometry-optimized *via* density functional theory. To ensure comprehensive coverage of the potential energy surface of the geopolymer, training data are generated over a temperature range spanning 300 K to 1000 K. The resulting first-principles forces and energies form the DP-GEN training dataset. Rigorous validation of the trained model is performed, followed by benchmarking against experimental measurements, classical MD results and DFT calculations.

Given the structural similarities between K-A-S-H and the N-A-S-H system, this work employs an innovative approach by leveraging a small amount of first-principles data from the K-A-S-H system to enable cross-system model transfer. The accuracy of the transferred model is then rigorously validated. The impact of this research lies in filling the critical gap of DP-GEN-based modeling in geopolymer systems, a field previously dominated by less accurate classical forcefields or computationally prohibitive AIMD. To the best of our knowledge, this is the first study to develop a DPMD model specifically tailored for both N-A-S-H and K-A-S-H geopolymer systems. The originality of this work lies in establishing a practical modeling strategy for complex amorphous geopolymer gels, where reliable potentials are constructed through system-specific dataset generation, AIMD sampling, and iterative refinement within a unified framework. By delivering a reliable and scalable simulation framework, the present study enables detailed atomistic understanding of the intrinsic links between the atomic structure and macroscopic properties in N-A-S-H and K-A-S-H, while also establishing a cost-effective transfer learning paradigm for future geopolymer research. The insights gained here are expected to inform bold, forward-looking strategies to guide the rational design and identification of advanced durable, high-performance geopolymer materials for infrastructure applications. For example, this forcefield can capture atomic scale deformation mechanisms to guide the development of materials with ultra-high mechanical strength, reliably simulate the behavior of geopolymers at 300–1000 K to optimize refractory formulations, and support systematic control of key element ratios such as the silicon aluminum ratio and sodium aluminum ratio, breaking free from the limitations of trial-and-error experiments.

2. Methods

2.1. Dataset establishment

In DP-GEN studies of geopolymers, preparing the dataset represents the essential first step. Constructing N-A-S-H molecular models and computing their energies and forces *via* first-principles methods lies at the heart of this stage. The process begins by building an N-A-S-H molecular model constructed from a sodium aluminosilicate (N-A-S) glass framework.³⁰ Because the amorphous structure of N-A-S glass closely mimics that of N-A-S-H gel, the glass model is first adjusted through selective atomic substitutions and charge balancing. A fraction of Si atoms is replaced with Al and free Ca atoms are introduced to mirror the typically high-calcium environment of geopolymers. This approach achieves the desired Si/Al molar ratio while maintaining overall charge neutrality and thus establishes the N-A-S backbone.³¹ Water molecules are then inserted to mimic gel hydration. A Monte Carlo (MC) scheme simulates water adsorption and diffusion until the required distribution of H₂O is reached. The resulting model is optimized at 500 K in a canonical ensemble (NVT) until a stable amorphous N-A-S-H network emerges.¹³

Once the amorphous N-A-S-H structure is generated, a deep optimization step *via* DFT refines the model. Specifically, after



structural relaxation and energy minimization, the initial training configurations are generated by AIMD simulations, in which atomic forces and stresses are directly computed from density functional theory and the temperature is explicitly controlled by the Nosé–Hoover thermostat. The initial AIMD sampling is performed at 500 K, which is selected as an intermediate temperature within the targeted 300–1000 K range to facilitate efficient subsequent DP-GEN exploration and convergence. Numerous iterative calculations adjust atomic positions and electronic densities to drive the system toward its lowest energy. This crucial stage not only ensures structural validity and stability but also reduces the computational demands of later first-principles runs. With the optimized structure in hand, the dataset generation phase employs fully *ab initio* quantum-mechanical calculations on the N-A-S-H system to extract atomic coordinates, energies and forces. Throughout the computational workflow, simulations run in an NVT ensemble governed by the Nose–Hoover thermostat that maintains temperatures between 300 K and 1000 K. The use of the NVT ensemble is based on the fact that the system volume has already been established during the preceding structural relaxation and energy minimization stage. In this step, the model is equilibrated by minimizing both atomic forces and stresses, with convergence thresholds of 0.01 eV Å⁻¹ for the force and 2.0 kBar for the stress, thereby ensuring a mechanically stable structure with a physically reasonable volume and density before finite-temperature sampling. Sampling across this wide range delivers comprehensive mapping of the potential energy surface and captures structures and energies under diverse thermodynamic states. In the subsequent DP-GEN concurrent-learning cycles, AIMD-based exploration is carried out at four temperatures, namely 300 K, 500 K, 700 K, and 1000 K. During each iteration, simulations at all temperatures are performed using identical sampling time and interval settings, while candidate configurations are selected according to the model deviation. This sampling procedure is adaptive, where temperature regions with larger uncertainty contribute more candidate structures, whereas fewer structures are retained from well-converged regions, ensuring both balanced thermodynamic coverage and efficient data generation. After integration and filtering, the initial dataset of 1230 atomic frames initiated 17 concurrent learning iterations of the DP-GEN framework, which integrated MD-based configuration space exploration. The initial dataset contains 1230 atomic configurations and mainly serves to initialize the DP-GEN concurrent-learning process. Through 17 iterative exploration–labeling–training cycles, the dataset is progressively expanded to a final total of 16 700 configurations. The adequacy of the training set is determined not by a predefined target number of frames, but by the convergence of model accuracy and the reduction of force deviation during active learning. In the final exploration stage, 137 834 out of 140 472 explored configurations fall within the predefined trust region, corresponding to a prediction accuracy of approximately 98%, which indicates a high level of model reliability. These richly detailed datasets support both training and validation of the DPMD model and pave the way for deep

understanding of how the microstructure of N-A-S-H geopolymers influences their macroscopic properties.

2.2. DP-GEN method

In this study, parallel learning-based model training is carried out within the DP-GEN framework, as schematically shown in Fig. 1. The workflow is organized into three sequential modules, including model training, configurational exploration, and data annotation. At the training stage, the DPMD model is trained with the deepmd-kit package on datasets generated either from initial training data or from previous iterations. The DPMD model begins by applying an embedding neural network to transform the initial atomic coordinates into distance-matrix representations that encode the relative distances between a reference atom and its neighboring atoms.^{28,32} The embedding network converts the raw local structural information, including atomic coordinates and interatomic distances within the cutoff radius, into symmetry-preserving descriptors. These descriptors are invariant with respect to translation, rotation, and permutation, and therefore provide a physically consistent representation of the local atomic environment suitable for machine learning. The fitting network then takes these descriptors as input and maps them to atomic energies. The total energy of the system is obtained by summing the atomic energy contributions over all atoms, while the atomic forces are derived as the gradients of the total energy with respect to atomic positions. In this way, the embedding and fitting networks work together to connect local atomic environments with the global energetic and force response of the system. These matrices are used to construct descriptors representing the local atomic environments. These descriptors are subsequently passed to the fitting network for the evaluation of atom-resolved energies and the corresponding forcefield. The cutoff radius is fixed at 7.0 Å, with a smoothing radius defined as 1.0 Å.²⁹ The embedding and fitting networks employ layer configurations of (30, 60, 120) and (300, 300, 300, 300), respectively. These network sizes are selected based on established practices in previous DPMD studies and are further examined in the present work through convergence tests.^{28,32,33} In particular, the relatively larger fitting network is adopted to capture the high-dimensional and highly nonlinear energy landscape of amorphous aluminosilicate systems, which involve multiple atomic species and strongly disordered local environments. This architecture provides a suitable balance between predictive accuracy and computational efficiency, whereas smaller network sizes lead to noticeable degradation in model performance. A plane-wave cutoff energy (ENCUT) of 550 eV is adopted, as shown in Fig. S1, where total energy convergence analysis confirms that it balances computational efficiency and numerical accuracy (see SI for details). The training of the potential energy surface essentially involves an optimization procedure for the parameters of both neural networks.³⁴ The network parameters are optimized through minimization of the loss function L , defined as,



$$L(p\varepsilon, pf, p\xi) = \frac{p\varepsilon}{N} \Delta\varepsilon^2 + \frac{pf}{3N} \sum_i |\Delta F_i|^2 + \frac{p\xi}{9N} \|\Delta\xi\|^2 \quad (1)$$

where N denotes the total number of atoms within the system, while $\Delta\varepsilon$, ΔF_i , and $\Delta\xi$ represent the discrepancies between the first-principles data and the predictions from the DPMD model for energy, atomic forces, and virial tensor, respectively. The coefficients $p\varepsilon$, pf , and $p\xi$ are predefined weighting factors that control the relative contributions of the energy, force, and virial terms in the loss function and are not treated as trainable variables. The actual trainable parameters of the model are the weights and biases of the neural networks, which are optimized during training to establish the mapping from atomic configurations to the corresponding energies and forces. Accordingly, the optimization of the loss function is carried out with respect to the neural-network parameters rather than the weighting coefficients themselves. The learning rate is with an initial value of 1.0×10^{-3} , followed by an exponential decay applied at intervals of 5000 training steps, eventually stabilizing at 5.0×10^{-8} . The optimization of the deep neural network model concludes after 5.0×10^5 steps.

During the exploration and labeling stages, deviation metrics are assessed based on an ensemble of previously trained models, with new configurations selected based on the maximum force deviation, defined as eqn (2),²⁸

$$\sigma_i^{\max} = \max_i \sqrt{\langle \|f_i - \langle f_i \rangle\|^2 \rangle} \quad (2)$$

where f_i represents the force acting on atom i and $\langle \|f_i - \langle f_i \rangle\|^2 \rangle$ corresponds to the ensemble-averaged value predicted by the DPMD model. Configurations exhibiting minor force deviations are highly likely to be sampled within the existing training data. In contrast, large force deviations imply that a configuration diverges markedly from physically meaningful trajectories, and such cases are excluded. Only configurations that fall within a predefined window are selected as candidates. This criterion is applied consistently across all sampled temperatures, and the retained candidate structures are those whose maximum force deviations indicate insufficient but still physically meaningful coverage by the current model. Practically, after executing multiple MD simulations, this criterion typically produces

hundreds to thousands of potential configurations. Because only a limited subset of these configurations is sufficient for improving the model performance, a cutoff-based selection strategy is therefore applied to narrow down the candidate set. The selected configurations are subsequently labeled and incorporated into the dataset for the following training iteration. While the labeling and training steps follow commonly used procedures, considerable flexibility exists in the sampling strategy for exploring the corresponding configuration space sampled at each iteration. A commonly adopted heuristic is to define the lower threshold of maximum deviation force marginally higher than the force prediction error of the trained model, while the upper bound is defined as the maximum deviation force $0.1\text{--}0.3 \text{ eV \AA}^{-1}$ larger than the lower bound.³⁵ In this paper, low and high maximum deviation forces are chosen as 0.12 eV \AA^{-1} and 0.25 eV \AA^{-1} , respectively, as shown by the gray vertical dashed line in Fig. 2.³⁵ These values are selected based on the force deviation distribution in Fig. 2, in which the lower bound aligns with the range where the model's prediction error stabilizes, while the upper bound (0.13 eV \AA^{-1} higher than the lower bound, falling within the $0.1\text{--}0.3 \text{ eV \AA}^{-1}$ interval) effectively excludes high-deviation outliers. This setup ensures that the training dataset retains most reliable configurations while balancing model accuracy and computational efficiency. The distribution of maximum force deviation for the N-A-S-H system spanning temperatures of 300 K, 500 K, 700 K, and 1000 K is presented in Fig. 2(a). Specifically, the distributions at all temperatures are concentrated in the low-deviation region, indicating that the developed deep-potential model accurately reflects the system's mechanical response. However, with increasing temperature, the distribution curves gradually broaden and shift slightly toward higher deviation values, reflecting the enhanced atomic thermal vibrations and increased local structural complexity at elevated temperatures, which impose greater demands on the fitting accuracy of the model. At 300 K, the distribution is the most concentrated, with nearly all snapshots exhibiting maximum force deviations within the low-value range, demonstrating excellent predictive stability of the model under low-temperature conditions. At 1000 K, although the distribution becomes broader, most data points remain below the threshold, suggesting that the model

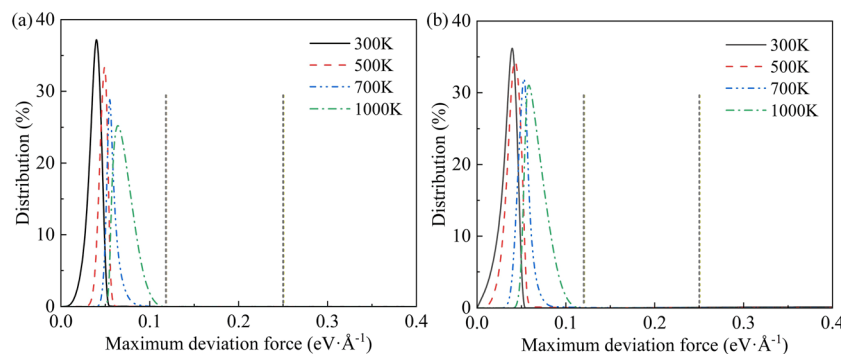


Fig. 2 Distributions of maximum deviation force for the (a) N-A-S-H and (b) K-A-S-H systems at different temperatures (300 K, 500 K, 700 K, and 1000 K), where the gray vertical dashed lines are the lower bound and higher bound of the selected maximum deviation force, respectively.



still maintains strong generalization capability at high temperatures. The results for N-A-S-H highlight the robustness of the DPMD model across different temperatures, while also revealing the challenges posed by structural complexity under elevated-temperature conditions to the accuracy of the model.

2.3. Transfer learning

To extend the applicability of the DPMD model initially developed for N-A-S-H to other polymeric systems such as K-A-S-H, a supplementary dataset is incorporated. Given that K-A-S-H and N-A-S-H share the same underlying aluminosilicate framework and differ solely in their exchangeable cation (Na^+ versus K^+), the construction of a new dataset from scratch is deemed unnecessary. Instead, fine-tuning is carried out using a limited set of first-principles calculations, with the optimized amorphous N-A-S-H configuration serving as the structural template for transfer.

The initial K-A-S-H structure is generated by substituting all Na^+ species within the optimized N-A-S-H network with K^+ counterparts. As both cations carry the same charge, the substitution preserved overall charge neutrality. The hydration environment established in N-A-S-H is preserved to approximate the gel chemistry of K-A-S-H. A similar equilibration protocol is applied, where equilibration is subsequently performed at 500 K in an NVT ensemble. This procedure yields a stable amorphous K-A-S-H configuration suitable for further refinement. DFT calculations are subsequently carried out under the same computational settings used for N-A-S-H, including the Perdew–Burke–Ernzerhof exchange–correlation functional, plane-wave cutoff energy, and total energy convergence threshold. NVT ensembles are conducted spanning temperatures from 300 K to 1000 K, with atomic coordinates, energies, and forces recorded at 5-fs intervals. From the resulting trajectories, only 330 frames are selected for training the DPMD model, significantly fewer than the 1230 frames used for N-A-S-H, thereby substantially reducing computational cost while maintaining model fidelity.

Afterwards, to capture the distinct structural features of K-A-S-H, including the coordination environment of K^+ ions and changes in network polymerization, the pre-trained N-A-S-H DPMD model is adapted rather than directly reused. Parameters associated with general aluminosilicate network features, including descriptors for $[\text{SiO}_4]$ and $[\text{AlO}_4]$ tetrahedral environments and energy modules for Si–O–Al bridging oxygens, are retained to serve as the foundation for transfer learning. Using the DP-GEN concurrent learning strategy, the initial 330 frames in the K-A-S-H DFT dataset forms the starting training set. Molecular trajectories are screened for configurations with force deviations exceeding $0.012 \text{ eV } \text{\AA}^{-1}$, resulting in 13 157 additional single-point DFT calculations. Fig. 2(b) presents the maximum force deviation distribution obtained for the K-A-S-H system, with its potential energy surface derived *via* transfer learning starting from the N-A-S-H model. Compared with N-A-S-H, the distributions of K-A-S-H at all temperatures are likewise concentrated in the low-deviation region, and their curve shapes closely resemble the corresponding profiles observed for

N-A-S-H, which demonstrates that the transfer-learning approach largely preserves the predictive strength of the parent model. At 300 K and 500 K, the distributions nearly coincide with those of N-A-S-H, suggesting that the transfer-learned model transitions smoothly to the K-A-S-H system under low-temperature conditions. At 700 K and 1000 K, although the distributions become slightly broader, the overall deviation levels remain low, and most snapshots still fall within a reasonable range. These results suggest that transfer learning achieves a substantial reduction in computational expense for new-system modeling while simultaneously preserving high predictive accuracy and stability in the complex configurational space under higher-temperature conditions. It is worth noting that, under high-temperature conditions, the K-A-S-H distributions exhibit greater smoothness relative to the corresponding N-A-S-H results, implying that transfer learning enhances, to some extent, the mode adaptability to different chemical environments. During training, the initial four layers of the fitting network are kept fixed, and only the parameters of the output layer are trained, and the training proceeds totally for 5×10^4 steps. A direct comparison between transfer learning and training from scratch for K-A-S-H is provided in the SI, where transfer learning is shown to achieve comparable accuracy with an approximately tenfold reduction in training steps. This transfer learning approach successfully adapts pre-trained aluminosilicate potential to a new cationic environment, providing a cost-efficient strategy for developing accurate interatomic potential across chemically similar systems.

2.4. Verification test

The trained DPMD model is subjected to a series of validation procedures to assess its reliability, demonstrating that both energies and atomic forces predicted within the DP-GEN framework closely reproduce DFT-level results across training and test datasets. It is noted that the data in the test set are not used for the training process. In the field of machine learning, root mean squared error (RMSE) serves as a standard metric for assessing model predictive performance.^{36,37} The mathematical definitions of the RMSE associated with energy and force are provided in eqn (3) and (4),

$$E_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{E}_i - E_i)^2} \quad (3)$$

$$F_{\text{RMSE}} = \sqrt{\frac{1}{n \cdot 3N} \sum_{i=1}^n \sum_{j=1}^3 \sum_{\alpha=1}^3 \left(\frac{\partial \hat{E}_i}{\partial r_{i,j\alpha}} - F_{i,j\alpha} \right)^2} \quad (4)$$

where n represents the total count of configurations contained in the dataset. N corresponds to the total number of atoms present in the system. E and F refer to the reference energies and forces obtained from DFT calculations. \hat{E} and \hat{F} indicate the corresponding energy and force values predicted by the DPMD model.

To further assess the performance of the DPMD model, lattice constants, radial distribution functions (RDFs), mean square displacement (MSD), and structural descriptors



characterizing geopolymer systems are computed. RDF profiles are obtained from canonical ensemble simulations conducted *via* AIMD and classical MD using various potential models. MSD analysis is performed to assess atomic mobility and dynamic stability, providing complementary insight into the model's ability to reproduce realistic transport behavior.^{38–40} For elastic property evaluation, the stress–strain method is applied with $\pm 5\%$ strain, which is a value that falls within the typical range for elastic property characterization and has been widely employed in simulations targeting elastic properties of geopolymers and similar materials.^{1,22,31,41} Elastic tensors under compressive deformation are evaluated using large-scale atomic/molecular massively parallel simulator (LAMMPS) for both DPMD and clay forcefield (ClayFF) models, while reference results based on DFT are taken from first-principles calculations. The bulk modulus, Young's modulus, shear modulus, and Poisson ratio are subsequently determined through Voigt–Reuss–Hill (VRH) averaging. These multi-faceted validations demonstrate the robustness and cross-system applicability of the DPMD model in capturing both static and dynamic properties of geopolymer systems.

3. Results and discussion

3.1. Evaluation of the DPMD model

The predictive capability of the DPMD model for the N-A-S-H system is evaluated by benchmarking its energy and force outputs against DFT results obtained for both training and validation configurations. Fig. 3 presents the corresponding energy–force correlation plots, where samples used for model fitting are distinguished from unseen test configurations by

different color markers. The reference diagonal indicates perfect agreement between model predictions and DFT values.

For the energy prediction, the points of both the training sets are tightly clustered around the diagonal line. This result demonstrates that the DPMD model is capable of faithfully recovering the DFT-derived energy values. The RMSE for energy in the training set is 0.003 eV per atom, and 0.005 eV per atom in the test set, both of which are fairly low compared to the literature²⁹ and have reached the level of meV deviation, demonstrating the high precision of the proposed model in energy prediction (see Fig. S2). Regarding the force prediction, the points also show a strong adherence to the diagonal line. The RMSE for force in the training set is 0.063 eV \AA^{-1} and 0.078 eV \AA^{-1} in the test set, as shown in Fig. S2. These small error values imply that the model can effectively capture the interatomic force information as accurately as DFT. In terms of computational efficiency, the DPMD model offers a substantial advantage over DFT. While a single DFT calculation for a N-A-S-H system of comparable size requires significant computational resources on a high-performance cluster, the trained DPMD model generates predictions almost instantaneously. The training process, though computationally intensive, is performed only once, after which the model can be reused for large-scale simulations or repeated evaluations with negligible additional cost.

Afterwards, the predictive performance of the transfer-learned DPMD model for the K-A-S-H system is evaluated by comparing its outputs with DFT reference values, as illustrated in Fig. 4. The model parameters are obtained *via* transfer learning from the previously trained N-A-S-H model. The four

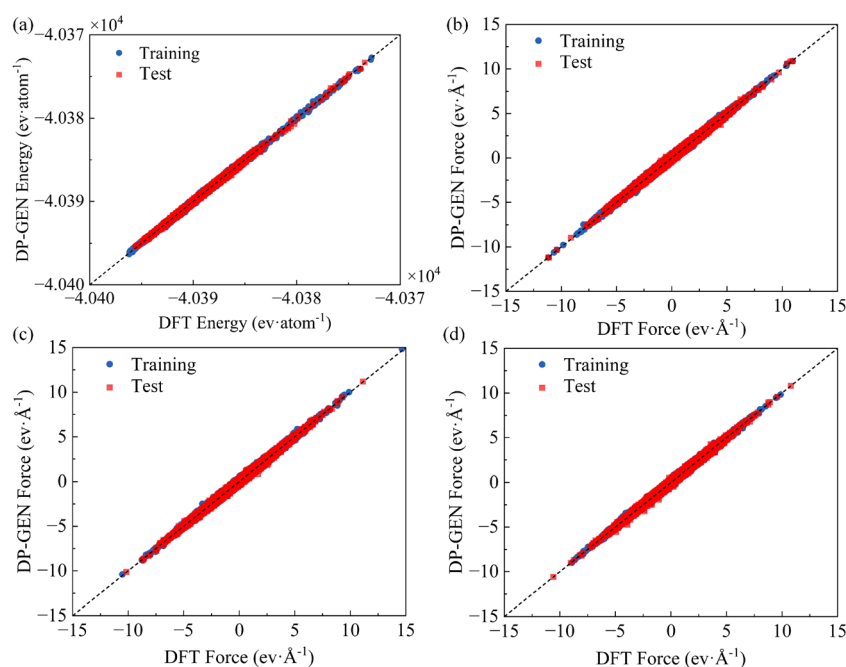


Fig. 3 Comparison with DFT on predicting the energy and force of N-A-S-H: (a) the results of total energy, and (b)–(d) show the atomic force along the *x*, *y*, and *z* directions, respectively. The close agreement across all panels demonstrates that the DPMD model reliably reproduces both energetic and force-related properties of the N-A-S-H system.



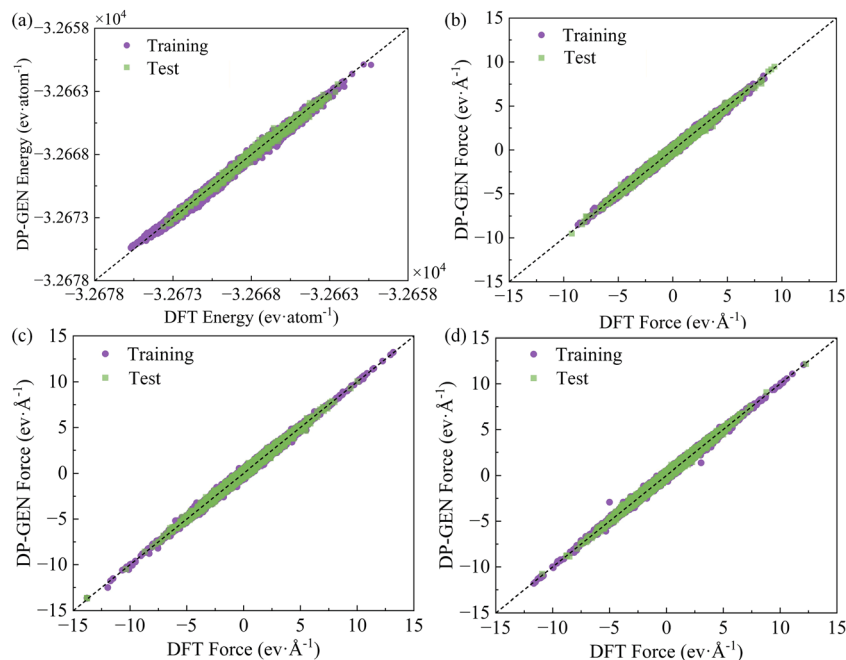


Fig. 4 Comparison with DFT on predicting the energy and force of K-A-S-H: (a) the results of total energy, and (b)–(d) show the atomic force along the x , y , and z directions, respectively. The close agreement across all panels demonstrates that the DPMD model reliably reproduces both energetic and force-related properties of the K-A-S-H system.

panels correspond to total energy, as shown in Fig. 4(a), and force components along the x , y , and z directions (see Fig. 4(b)–(d)), with green points representing the training set and purple points, the test set. The dashed diagonal line in each plot denotes perfect agreement with DFT. Across all subplots, data points for both training and test sets are densely distributed along the diagonal, indicating that the transfer-learned model maintains high predictive fidelity for both energy and atomic forces in the new chemical system. For energy prediction, the clustering of points is particularly tight, reflecting minimal deviation from DFT values. For force components, strong linear correlations are also observed in all three Cartesian directions, with only slight dispersion relative to the diagonal, consistent with the inherent sensitivity of force calculations to local atomic environments. The close agreement between predictions and DFT results demonstrates that the structural and energetic features learned from N-A-S-H can be effectively generalized to K-A-S-H *via* transfer learning. This approach significantly reduces the amount of new DFT data and training time required, while still achieving DFT-level accuracy.

When compared with the outcomes obtained for the N-A-S-H system, the K-A-S-H model exhibits comparable RMSE values for both energy and force predictions. The RMSE for energy is 0.002 eV per atom and 0.003 eV per atom in the training and test set, respectively, and the RMSE for force is 0.090 eV \AA^{-1} and 0.092 eV \AA^{-1} in the training and test set, respectively, confirming that transfer learning preserves the accuracy of the original model while enabling rapid adaptation to a chemically related system, as shown in Fig. S3. This highlights the efficiency and scalability

of the DP-GEN framework for extending interatomic potentials across similar material systems.

3.2. Structural properties of N-A-S-H and K-A-S-H

3.2.1 RDF analysis. A reasonable and stable structural model is always essential for initiating MD simulations. After verifying the accuracy of the constructed N-A-S-H structure, the reliability of the DPMD model in describing interatomic interactions within N-A-S-H is further assessed by calculating and comparing the RDF of different atomic pairs at 300 K, as shown in Fig. 5. It is found that the results obtained from DPMD exhibit a high degree of consistency with AIMD in both peak positions and overall curve profiles, whereas ClayFF shows significant deviations. In the Na–O, Si–O, and H–O subfigures, the first peak heights predicted by ClayFF are generally higher than those from DPMD and AIMD, indicating an overestimation of local ordering for these atomic pairs. Such an overestimation may lead to inaccurate evaluations of coordination numbers and local structural stability.^{13,20} In the Al–O case, ClayFF produces a pronounced shift in the first peak position, further demonstrating its limited accuracy in capturing the Al–O coordination environment within the tetrahedral framework, as shown in Fig. 5(b). By contrast, the peak positions from DPMD nearly coincide with those from AIMD, although the overall peak intensities are slightly lower. More quantitatively, for the N-A-S-H system, the first-peak position differences between DPMD and AIMD are generally within about 0.02 \AA for the framework pairs (Si–O and Al–O), while the corresponding peak-height deviations are typically on the order of 10%. For the more weakly bound Na–O and H–O pairs, the peak-position



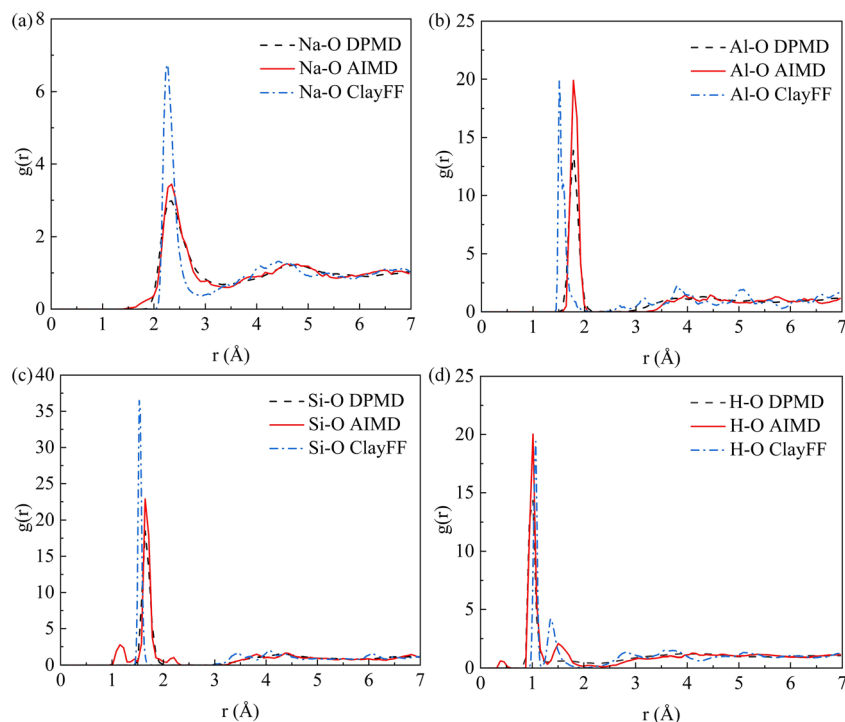


Fig. 5 RDFs of N-A-S-H for (a) Na–O, (b) Al–O, (c) Si–O, and (d) H–O atomic pairs at 300 K, comparing results from DPMD, AIMD, and ClayFF to quantify the consistency of DPMD with ab initio-level AIMD data and its superiority over the classical ClayFF potential, and verifying the high accuracy of DPMD in describing short-range atomic correlations of the N-A-S-H system.

differences are slightly larger, typically within 0.04–0.06 Å, and the peak-height deviations are generally within about 6–8% at 300 K. These results indicate that DPMD reproduces the primary coordination shell with high fidelity, whereas the remaining differences mainly arise from a slight broadening of the local structural distribution. This discrepancy may stem from a minor underestimation of local ordering during the fitting process of the DPMD model. Nevertheless, this difference does not compromise the reproduction of the overall structural features, as DPMD still accurately captures the primary coordination shell information.

It is important to emphasize that RDF curves reflect the probability distribution of specific atomic pairs at different distances, and the peak positions correspond to average bond lengths or coordination shell locations, while the peak heights represent the degree of local ordering and the strength of interatomic interactions.⁴² Therefore, the strong agreement between DPMD and AIMD across various atomic pairs not only confirms that the model can reliably forecast both energy and force behaviors but also demonstrates its capability to faithfully reproduce the local structural characteristics and chemical environments associated with the N-A-S-H material system. In contrast, the deviations observed in ClayFF highlight the limitations of traditional empirical forcefields in complex inorganic polymer systems, particularly in cases involving multiple components and diverse coordination environments. The RDFs of N-A-S-H at other temperatures (500 K, 700 K and 1000 K) are shown in Fig. S4–S6, where Si–O and Al–O retain short-range coordination stability but show broadened peaks and reduced

intensity with increasing temperature, while Na–O and H–O exhibit more pronounced peak attenuation and broadening, accompanied by progressive deterioration of medium-range order and enhanced ionic mobility. In summary, the developed DPMD model achieves structural prediction accuracy comparable to AIMD while offering significant computational efficiency, thus providing a reliable tool for subsequent investigations into the structure and properties of geopolymer systems. From a mechanistic perspective, the RDF results also indicate that charge-balancing cations play an important role in regulating framework compactness and rigidity. In the N-A-S-H system, the relatively localized Na–O coordination environment suggests stronger electrostatic interactions between Na⁺ and framework oxygen atoms, which help stabilize the aluminosilicate network and promote a more compact local structure. Such stronger cation-framework association is expected to enhance network constraint and reduce structural flexibility, thereby contributing to the comparatively higher rigidity of N-A-S-H.

Afterwards, to examine the applicability of the transfer learning strategy in the K-A-S-H system, the RDFs between different atoms at 300 K are further analyzed, as shown in Fig. 6. The results from DPMD are generally consistent with AIMD in terms of peak positions, successfully reproducing the structural features of the primary coordination shells, whereas ClayFF exhibits noticeable deviations. For the K–O, Si–O, and H–O pairs, the peak intensities obtained from ClayFF are significantly higher than those from DPMD and AIMD, indicating an overestimation of local ordering. In the case of the Al–O pair,



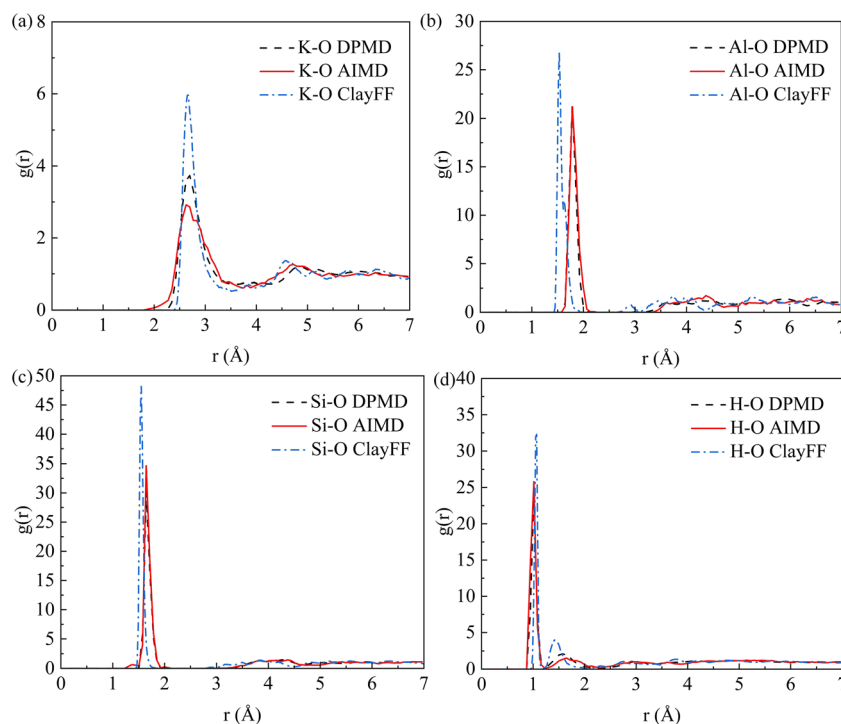


Fig. 6 RDFs of K-A-S-H for (a) K–O, (b) Al–O, (c) Si–O, and (d) H–O atomic pairs, comparing results from DPMD, AIMD, and ClayFF to quantify the consistency of DPMD with ab initio-level AIMD data and its superiority over the classical ClayFF potential, and verifying the high accuracy of DPMD in describing short-range atomic correlations of the K-A-S-H system.

the first peak position predicted by ClayFF is shifted, reflecting its limitations in accurately describing the tetrahedral framework. By contrast, the curve profiles and peak intensities from DPMD show excellent agreement with AIMD, demonstrating that the K-A-S-H DPMD potential derived through transfer learning from the N-A-S-H model maintains high predictive accuracy in a new cationic environment. Quantitatively, the first-peak position differences between DPMD and AIMD for K-A-S-H remain within approximately 0.02–0.03 Å for Si–O and Al–O, and within about 0.05–0.06 Å for K–O and H–O. The corresponding peak-height differences are generally within 5% for the framework pairs at 300 K. This confirms that the transfer-learned model preserves the local structural characteristics of the parent system with only limited deviations in peak sharpness and local ordering. This approach not only retains the broad generalization ability of the original model, but also enhances the overall computational efficiency compared with AIMD. In addition to validating the transfer learning strategy, the RDF comparison also reveals the structural role of K^+ in the geopolymer network. Compared with Na^+ , K^+ exhibits a broader and less localized first coordination shell with oxygen, which is consistent with its larger ionic radius, lower charge density, and weaker electrostatic binding to the aluminosilicate framework. This more delocalized coordination environment implies a looser local packing state and a broader distribution of cation-oxygen distances, thereby increasing medium-range structural variability and reducing the overall rigidity of the K-A-S-H network. These results further confirm that the transfer learning method offers strong reliability and broad applicability

for cross-system modeling of complex geopolymer systems. The RDFs of K-A-S-H at 500 K, 700 K and 1000 K are shown in Fig. S7–S9, where the faster deterioration of the medium-range structure of ionic pairs at elevated temperatures is observed, which further validates the model's capability to capture temperature-dependent structural dynamic behaviors.

3.2.2 MSD analysis. The MSDs of the N-A-S-H system as a function of time at 300 K, obtained from DPMD, AIMD, and ClayFF simulations are shown in Fig. 7. The overall trend indicates that hydrogen atoms exhibit the highest mobility, with their MSD increasing rapidly over time and reaching values far greater than those of other species, which aligns well with findings reported in prior molecular-level simulations.⁴³ Sodium ions show the next highest mobility, while the MSD characteristics associated with Si, Al, and O atoms remain consistently low, reflecting the stability of the framework atoms and their localized vibrational motion.⁴¹ This observation is consistent with the structural characteristics of geopolymers, in which at room temperature the framework atoms primarily maintain network stability, whereas small-radius ions and light elements display stronger migration within pores or channels.⁴⁴ The MSD results obtained from different simulation methods exhibit consistent overall trends but reveal notable quantitative discrepancies. For Na and H, the results from DPMD and AIMD are in excellent agreement, with nearly identical magnitudes and growth rates, confirming that the DPMD model accurately reproduces the ionic dynamics captured by first-principles simulations. More specifically, for the mobile ions in N-A-S-H at 300 K, the MSD difference between DPMD and AIMD remains



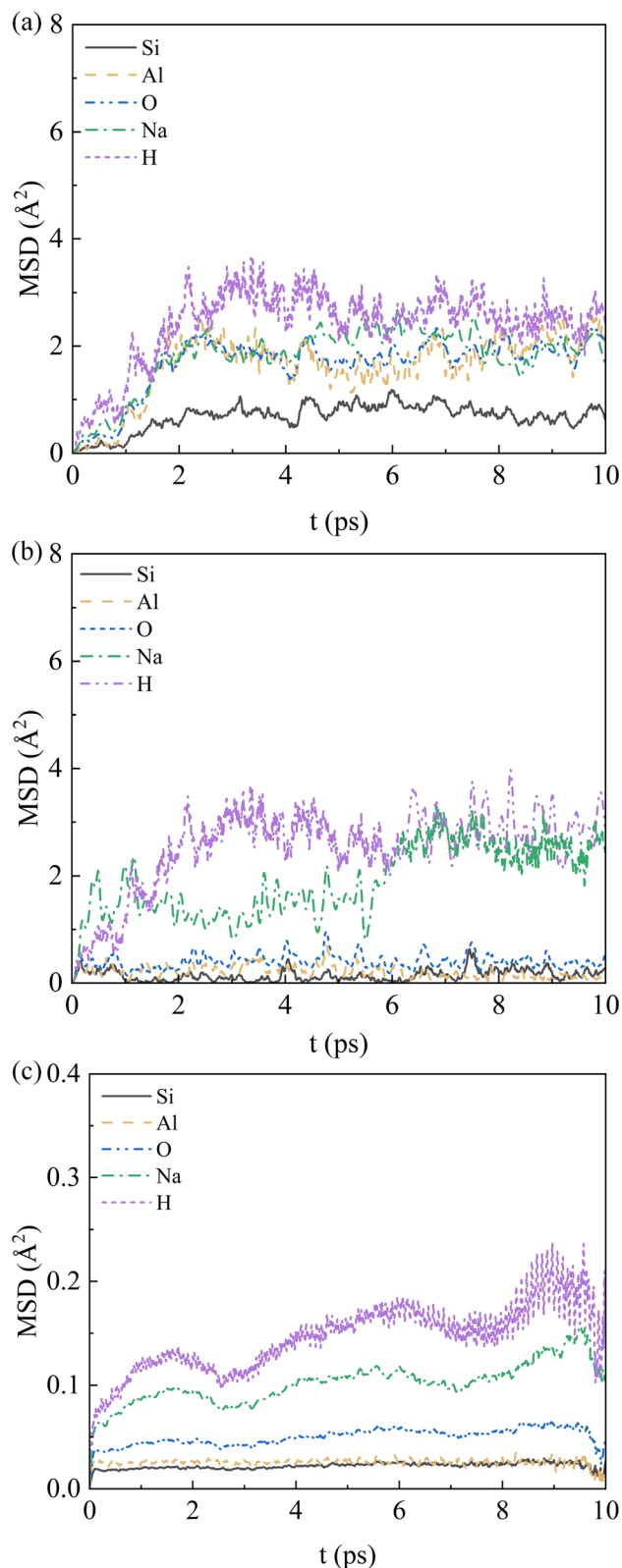


Fig. 7 MSD of the N-A-S-H system obtained from (a) DPMD, (b) AIMD, and (c) ClayFF simulations at 300 K. The DPMD-calculated MSD curves show excellent quantitative agreement with *ab initio* AIMD results for both framework atoms and mobile ions, whereas ClayFF exhibits noticeable systematic deviations in ionic diffusion coefficient predictions, highlighting the superiority of DPMD in balancing accuracy and efficiency for probing ionic dynamics and structural stability.

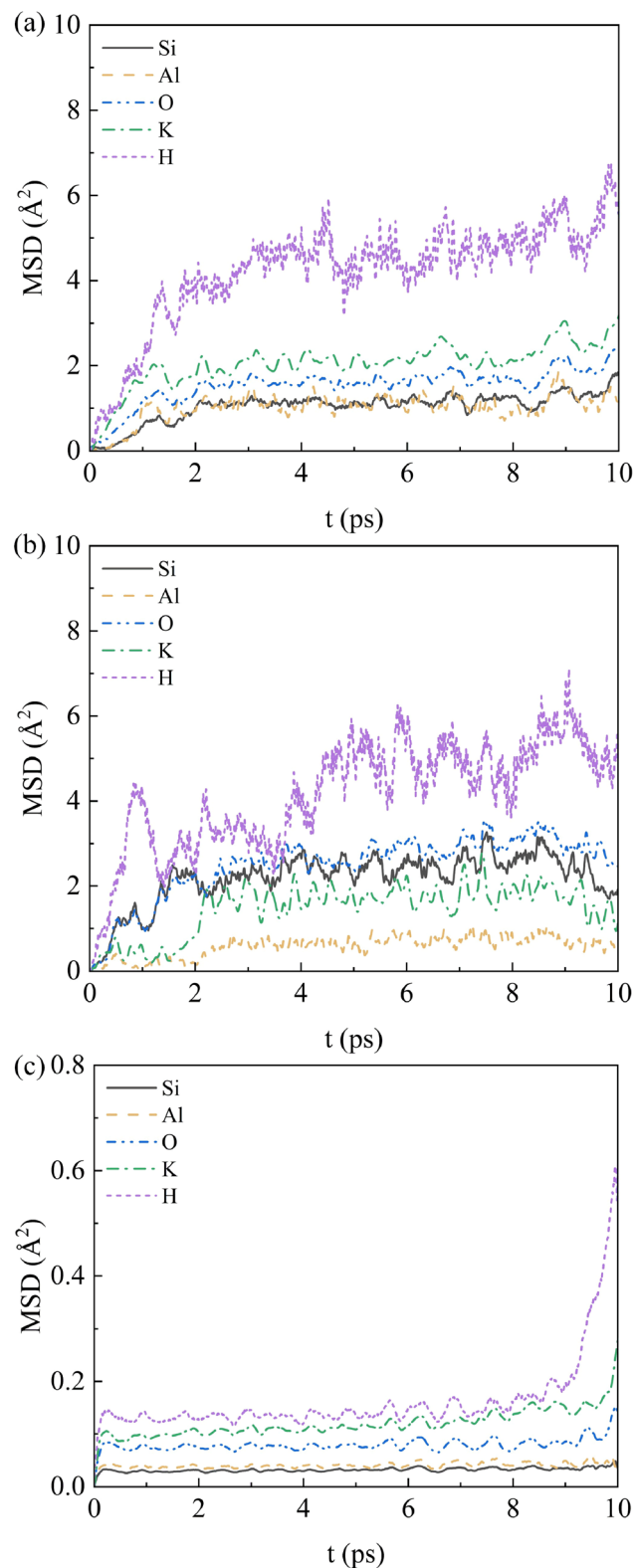


Fig. 8 MSD of the K-A-S-H system obtained from (a) DPMD, (b) AIMD, and (c) ClayFF simulations at 300 K. The DPMD-calculated MSD curves show excellent quantitative agreement with *ab initio* AIMD results for both framework atoms and mobile ions, whereas ClayFF exhibits noticeable systematic deviations in ionic diffusion coefficient predictions, highlighting the superiority of DPMD in balancing accuracy and efficiency for probing ionic dynamics and structural stability.



approximately 8% over the main diffusion regime. For the framework atoms, the relative deviations are larger because their absolute displacements are much smaller and therefore more sensitive to local vibrational details. In terms of absolute values, the end-of-trajectory MSD differences are generally within about 0.5–2.0 Å² for Si and Al, and within 1–3 Å² for O, indicating that the deviation is still limited in magnitude despite the higher relative sensitivity of the framework atoms. For the framework atoms (Si, Al, and O), the MSD values predicted by DPMD are slightly higher than those from AIMD, which can be attributed to the smoother representation of short-range repulsion and vibrational amplitudes in the machine-learning potential, leading to marginally enhanced thermal motion of the network. In contrast, the results from ClayFF deviate by orders of magnitude from both DPMD and AIMD, significantly overestimating the mobility of ions, particularly H. This discrepancy arises from the fixed-charge, non-polarizable nature of the classical forcefield, which fails to capture the directionality and strength of hydrogen bonding as well as the correct energy barriers for proton and cation migration. The diffusion of Na is also overestimated, although to a lesser extent than H, reflecting the limitations of the simplified electrostatic and Lennard–Jones interactions in describing the complex coordination environment of charge-balancing cations. Meanwhile, the MSD of the N-A-S-H system at other temperatures is shown in Fig. S10, in which framework atoms (Si, Al, and O) maintain low MSD with slight temperature-dependent increases and mobile ions (Na and H) show enhanced mobility. Overall, DPMD achieves a high level of consistency with AIMD, whereas ClayFF, despite capturing the qualitative trends, exhibits systematic deviations in the quantitative prediction of ionic diffusion. This comparison highlights the advantage of DPMD models for investigating ionic dynamics and structural stability in complex inorganic polymer systems. Mechanistically, the relatively lower mobility observed in N-A-S-H can be associated with the stronger interaction between Na⁺ and the negatively charged framework oxygen atoms. This stronger cation-framework coupling imposes greater local constraint on both ion transport and framework relaxation, thereby limiting large-scale structural rearrangements and maintaining a comparatively rigid network at finite temperature. Unlike the empirical potential function ClayFF, which relies on fixed pairwise interaction parameters, DPMD is trained on high-precision *ab initio* data and can accurately capture the temperature-dependent weak interactions between mobile ions and the rigid silicoaluminate framework. These interactions directly determine the slope of the MSD curve, a core indicator of diffusion ability, and the low-MSD plateau of framework atoms, a characteristic manifestation of structural stability—a key reason why ClayFF exhibits systematic biases in the quantitative prediction of ion diffusion coefficients.

Similarly, the MSD of the K-A-S-H model is analyzed using different methods at 300 K, as shown in Fig. 8. At 300 K, the MSD profiles of the K-A-S-H system display marked dynamical distinctions relative to those of N-A-S-H. The MSD values of K-A-S-H are consistently higher than those of N-A-S-H, for both framework atoms (Si, Al and O) and mobile species (K and H),

indicating stronger thermal vibrations and enhanced ionic mobility. This phenomenon primarily arises from the larger ionic radius and reduced charge density of K⁺ compared with Na⁺, which weakens its electrostatic interaction with framework oxygen atoms. Therefore, K⁺ is less tightly confined within the aluminosilicate network and experiences a lower effective migration barrier. At the same time, the weaker cation–framework interaction reduces the topological constraint exerted on the surrounding framework, allowing larger-amplitude thermal fluctuations and more pronounced local structural rearrangements. Therefore, the stronger dynamics in K-A-S-H originate not only from the intrinsic mobility of K⁺ itself, but also from the increased flexibility of the entire cation-framework coupled network. As a result, local structural relaxation is intensified, the thermal amplitude of the framework atoms is enlarged, and more accessible diffusion pathways are created for ionic transport. Consequently, K-A-S-H exhibits higher overall mobility than N-A-S-H under ambient conditions. In terms of methodological comparison, the results of DPMD and AIMD for K-A-S-H remain in excellent agreement, and the MSD curves of both framework and ionic species nearly overlap in magnitude and growth rate, confirming the robustness and cross-system transferability of the K-A-S-H DPMD model obtained through transfer learning based on the N-A-S-H potential. Quantitatively, for the mobile species in K-A-S-H at 300 K, the MSD difference between DPMD and AIMD remains within about 10% over the main diffusion regime. As in N-A-S-H, the framework atoms exhibit larger relative deviations because of their smaller absolute displacements, while the end-of-trajectory absolute differences remain limited, typically within about 0.5–2.0 Å² for Si and Al, and up to 1–3 Å² for O. This result indicates that DPMD models not only achieve near-DFT accuracy within a single system but can also be effectively extended to chemically related systems through transfer learning, thereby capturing subtle dynamical differences. By contrast, the results from ClayFF are significantly lower than those of DPMD and AIMD, with particularly pronounced underestimation of the MSD of K and H, highlighting the systematic limitations of classical forcefields in describing hydrogen bonding, proton dynamics, and the interactions between large alkali cations and the aluminosilicate framework. The MSDs of the K-A-S-H model at other temperatures are shown in Fig. S11, where atomic mobility increases markedly with temperature.

The close consistency between DPMD and AIMD not only confirms the accuracy of DPMD models but also underscores the methodological advantage of transfer learning in modeling complex inorganic polymer systems. In contrast, although ClayFF captures qualitative trends, it systematically underestimates quantitative behavior and fails to reproduce the true dynamics of N-A-S-H and K-A-S-H. Therefore, transfer-learning-based DPMD emerges as a powerful and efficient approach that combines high accuracy with computational scalability, making it an ideal tool for investigating the dynamical properties of complex inorganic polymers.

3.2.3 Local structural order parameter analysis. While the RDF and MSD analyses provide valuable insights into the thermally induced changes in atomic pair correlations and



long-range mobility, they do not fully capture the geometric integrity and local symmetry of the coordination environments. To further elucidate the structural evolution at the atomic scale, especially the distortion and rearrangement of polyhedral units under thermal and dehydration effects, we introduce the local structural order parameter q as a complementary metric.^{45,46} This parameter enables a more nuanced characterization of the short-range structural motifs, allowing us to quantify deviations from ideal configurations such as tetrahedral, octahedral, and defective coordination states.⁴⁷ The q parameter quantifies the similarity between local atomic arrangements and ideal geometries, with higher values indicating greater structural order. Its mathematical definition is provided below:

$$q = 1 - \frac{3}{8} \sum_{k>i} \left(\frac{1}{3} - \cos \theta_{ijk} \right)^2 \quad (5)$$

where θ_{ijk} is the bond angle formed by the central atom j and its two adjacent atoms i and k , and the q value is calculated by statistically analyzing the bond angles of all adjacent atom pairs.

The distributions of the q -order parameter for the N-A-S-H system at 300 K calculated using different models are shown in Fig. 9, aiming to characterize the evolution of various types of coordination, including four-fold, five-fold, tetrahedral, and octahedral configurations, under thermal perturbation. Vertical dashed lines mark the reference q values corresponding to ideal configurations, facilitating the identification of deviations at different temperatures. The distribution results reveal a high degree of consistency between DPMD and AIMD in terms of peak positions, profile shapes, and the relative proportions of various coordination motifs. This agreement confirms that the DPMD model achieves structural resolution comparable to that of first-principles methods, demonstrating reliable physical accuracy. More quantitatively, the tetrahedral peak remains centered near $q = 1.0$ in both DPMD and AIMD, with a peak-position difference typically smaller than about 0.02–0.04. The neighboring defective or distorted motifs in the q range of approximately 0.3–0.9 also show only limited peak-position differences, generally below 0.2. The main residual discrepancy lies in the relative intensities of low-symmetry and defective local motifs, which indicates that DPMD captures the dominant local geometry accurately while retaining only limited statistical deviation in the distribution of distorted environments. At 300 K, the overall distribution is narrow and concentrated, with sharp peaks near the ideal values for tetrahedral and three-fold structures. This indicates that Si–O and Al–O tetrahedral units maintain well-defined geometries, serving as the primary contributors to structural stability at low temperature. The fractions of five-fold and octahedral configurations are small, suggesting that the system is predominantly composed of ideal tetrahedral units with high local order.^{48–50} In contrast, the ClayFF simulation exhibits pronounced deviations across multiple coordination types, particularly with anomalous differences in low-coordination regions. Additionally, the tetrahedral peak in ClayFF is both shifted and diminished in intensity, indicating a reduced ability to capture the correct

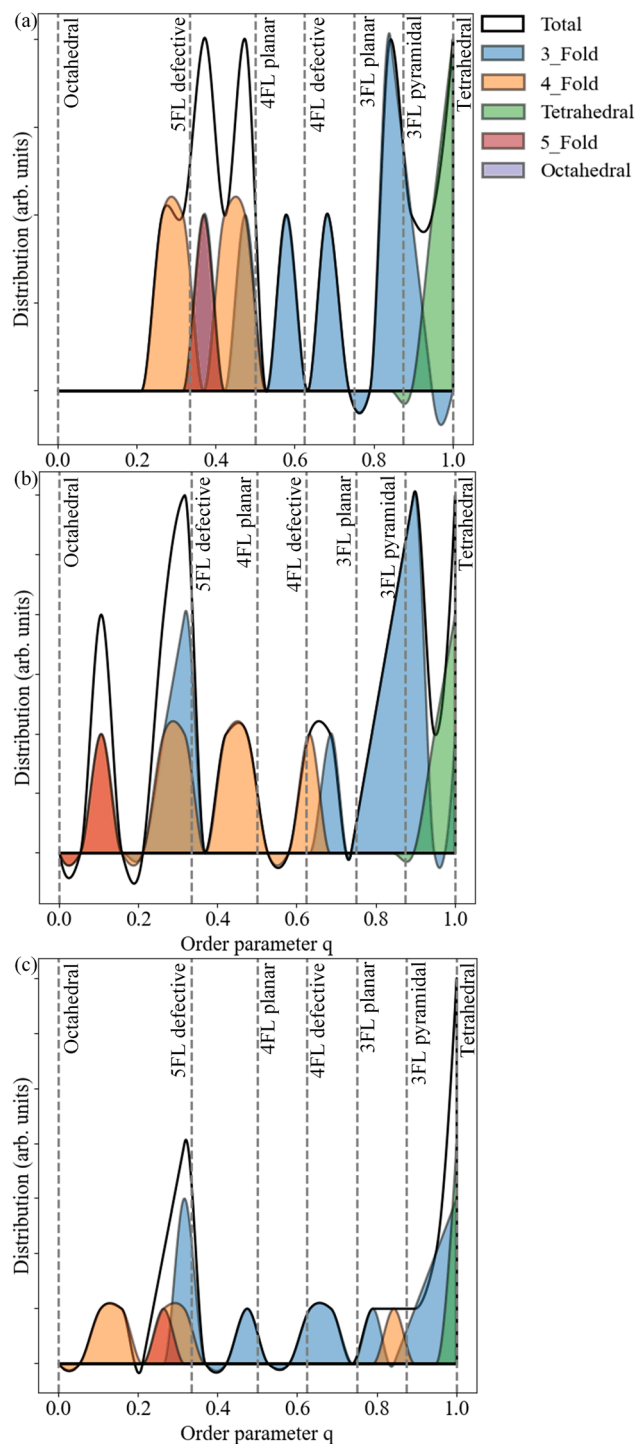


Fig. 9 q -order parameters of the N-A-S-H system obtained from (a) DPMD, (b) AIMD, and (c) ClayFF simulations at 300 K. The vertical lines indicate ideal local order values. The DPMD-calculated q -order parameters show excellent consistency with AIMD results in both peak positions and peak intensities, whereas ClayFF predictions exhibit noticeable deviations from both the ideal values and AIMD benchmarks in these two key aspects. This confirms the superior accuracy of DPMD in characterizing the local atomic ordering of the N-A-S-H system compared to the classical ClayFF potential.



local geometry. These discrepancies suggest inherent limitations in classical forcefields when describing the fine structural features of geopolymer systems, likely stemming from their simplified treatment of angular potentials and electrostatic interactions. Meanwhile, the q -order parameters of the N-A-S-H system at other temperatures are shown in Fig. S12, where tetrahedral structure distributions broaden with decreasing peak heights as temperature increases, and five-fold and low-coordination structures increase. These results indicate that the N-A-S-H framework remains dominated by tetrahedral local order at low temperature, while thermal activation mainly induces progressive local distortion rather than immediate collapse of the framework. This behavior is consistent with the RDF and MSD analyses, which show that the aluminosilicate network retains comparatively strong structural constraint in the presence of Na^+ .

Afterwards, the distributions of q -order parameters for the K-A-S-H system at 300 K from different models are presented in Fig. 10. In comparison with the N-A-S-H system, the framework cation in K-A-S-H is K^+ , whose larger ionic radius and weaker electrostatic interactions result in a more pronounced structural response under thermal perturbation.^{51–53} At 300 K, the distributions of four-fold and tetrahedral configurations are sharply peaked near their ideal q values, indicating that the system is predominantly composed of well-defined Si–O and Al–O tetrahedral units exhibiting pronounced local ordering. The fraction of three-fold and five-fold coordination is limited, likely associated with boundary regions or oxygen-deficient sites. Octahedral configurations are nearly absent, suggesting that the framework remains structurally stable at low temperature without significant coordination transitions. In both the DPMD and AIMD results, the tetrahedral configuration exhibits a distinct peak, suggesting that the Si–O and Al–O tetrahedral units in the K-A-S-H network preserve substantial geometric ordering at ambient temperature. Moreover, the distributions in the low- and high-coordination regions are largely consistent between the two methods, with no significant anomalies observed, suggesting that DPMD accurately reproduces the local structural features captured by AIMD. Quantitatively, the main tetrahedral-related peak positions remain nearly unchanged between DPMD and AIMD, with deviations generally smaller than about 0.04. In the distorted or defective coordination regions, the peak-position differences are also limited, typically below 0.2, whereas the relative intensity deviations are somewhat larger, particularly for five-fold and low-symmetry motifs. This indicates that the transfer-learned DPMD model accurately reproduces the dominant local structural motifs while still showing modest statistical differences in the populations of highly distorted environments. It is noteworthy that, although the K^+ ion possesses a relatively large ionic radius and weak electrostatic binding, which could theoretically promote framework relaxation and distortion, neither DPMD nor AIMD reveals substantial coordination transitions or structural rearrangements at 300 K. This indicates that the system maintains good structural stability under low-temperature conditions. This behavior contrasts with that observed in the N-A-S-H system, which exhibits stronger tetrahedral dominance and

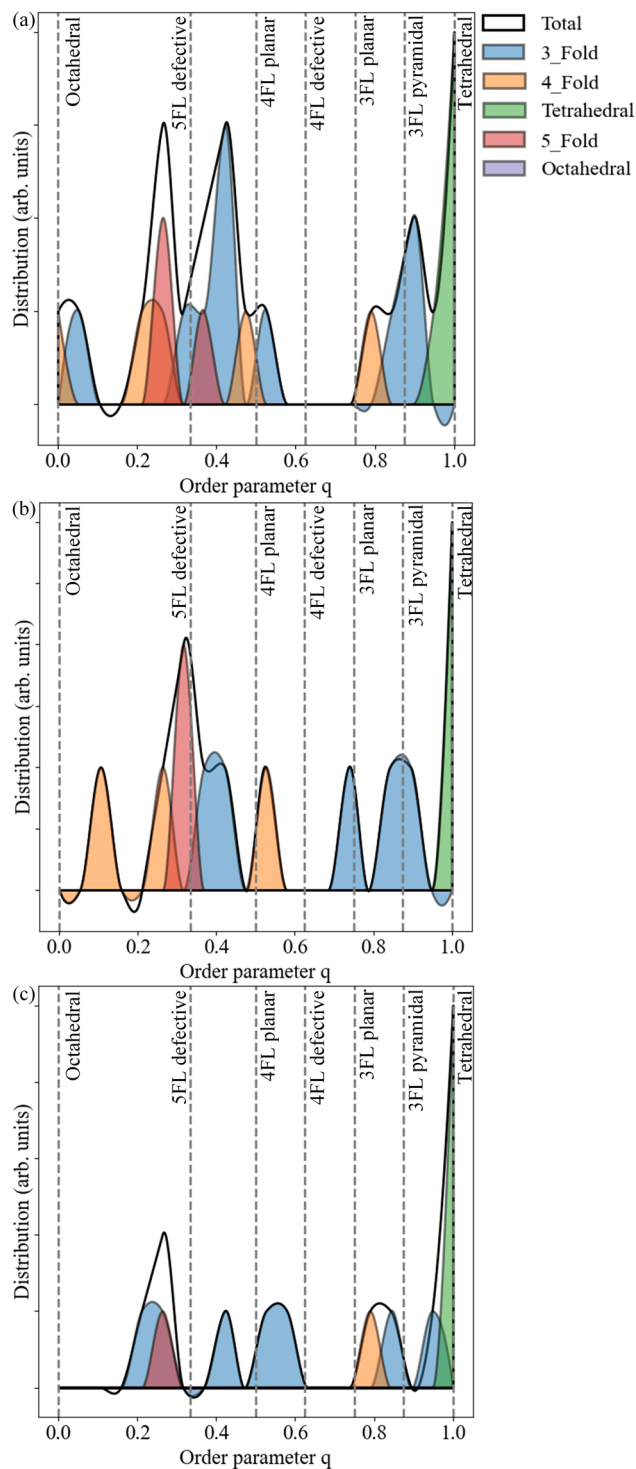


Fig. 10 q -order parameters from the K-A-S-H system obtained from (a) DPMD, (b) AIMD, and (c) ClayFF simulations at 300 K. The vertical lines indicate ideal local order values. The DPMD-calculated q -order parameters show excellent consistency with AIMD results in both peak positions and peak intensities, whereas ClayFF predictions exhibit noticeable deviations from both the ideal values and AIMD benchmarks in these two key aspects. This confirms the superior accuracy of DPMD in characterizing the local atomic ordering of the K-A-S-H system compared to the classical ClayFF potential.



lower coordination diversity at the same temperature. In comparison, the ClayFF simulation yields markedly different distributions across several coordination types, with pronounced differences in the three-fold and four-fold regions, and noticeable shifts in both the position and intensity of the tetrahedral peak. These discrepancies highlight the limitations of classical forcefields in accurately describing the local geometry of the K-A-S-H framework, likely due to simplified treatments of ionic size effects and angular potentials. Therefore, DPMD demonstrates a high level of agreement with AIMD in terms of structural resolution, effectively reproducing the local coordination distributions observed in first-principles simulations. This consistency validates the reliability of DPMD as an efficient surrogate for AIMD in modeling the K-A-S-H system. The q -order parameters of the K-A-S-H model at other temperatures are shown in Fig. S13, in which tetrahedral distributions broaden with decreasing peak intensity as temperature increases, and five-fold and octahedral configurations increase significantly (more pronounced than in N-A-S-H). Compared with N-A-S-H, the more pronounced broadening of the tetrahedral-related distributions and the stronger increase in distorted coordination motifs suggest that the K-A-S-H framework is more susceptible to local rearrangement. This provides complementary structural evidence that the weaker K^+ -framework interaction leads to a more flexible network, which is consistent with the enhanced dynamics and lower stiffness observed for K-A-S-H.

In summary, the strong agreement between DPMD and AIMD validates the feasibility and reliability of DPMD as an efficient surrogate for *ab initio* simulations in geopolymer modeling. The deviations observed in ClayFF further underscore the advantages of DPMD in chemically complex systems, providing a robust foundation for large-scale and long-time-scale simulations of geopolymers.

3.3. Density and mechanical properties of N-A-S-H and K-A-S-H

To further assess the reliability of the developed DPMD model, a set of mechanical properties of N-A-S-H, namely density, bulk modulus, Young's modulus, shear modulus, and Poisson ratio, are systematically quantified and benchmarked against available experimental measurements as well as results from conventional molecular simulations, as summarized in Table 1.^{54–57} The DPMD-predicted density of the N-A-S-H system is $2.06 \pm 0.02 \text{ g cm}^{-3}$, which falls within the experimental range of $1.98 \pm 0.12 \text{ g cm}^{-3}$, whereas the value obtained from ClayFF (2.91 g cm^{-3}) is significantly overestimated, indicating an overestimation of structural densification due to excessive constraints on local relaxation and porosity. For the mechanical properties, the bulk modulus (measured to be $36.10 \pm 2.55 \text{ GPa}$), Young's modulus (measured to be $36.46 \pm 2.37 \text{ GPa}$), and shear modulus (measured to be $13.69 \pm 1.18 \text{ GPa}$) predicted by DPMD fall close to the upper bounds of the experimental ranges (bulk modulus $16.32 \pm 4.15 \text{ GPa}$, Young's modulus $21.31 \pm 5.24 \text{ GPa}$, and shear modulus $8.58 \pm 1.62 \text{ GPa}$).^{54–57} This indicates that DPMD not only maintains consistency with experimental

Table 1 The mechanical properties of N-A-S-H obtained from DPMD, ClayFF, and the experiment, respectively

Properties	DPMD	ClayFF	Experiment ^{54–57}
Density (g cm^{-3})	2.06 ± 0.02	2.91 ± 0.05	1.98 ± 0.12
Bulk modulus (GPa)	36.10 ± 2.55	100.19 ± 6.16	16.32 ± 4.15
Young's modulus (GPa)	36.46 ± 2.37	118.59 ± 5.20	21.31 ± 5.24
Shear modulus (GPa)	13.69 ± 1.18	47.58 ± 3.78	8.58 ± 1.62
Poisson ratio	0.33 ± 0.04	0.31 ± 0.06	0.26 ± 0.02

observations but also offers a credible representation of the stiffness and deformation resistance of the material. In contrast, ClayFF significantly overestimates these properties, with values of $100.19 \pm 6.16 \text{ GPa}$, $118.59 \pm 5.20 \text{ GPa}$, and $47.58 \pm 3.78 \text{ GPa}$, respectively, which exaggerate the mechanical strength of the system and highlight its limitations in capturing the flexibility and structural relaxation of inorganic polymers. Regarding the Poisson ratio, the DPMD result (0.33 ± 0.04) is slightly higher than the experimental range 0.26 ± 0.02 but still within a reasonable margin, providing a reliable description of the lateral deformation behavior under stress. Although the ClayFF value (0.31 ± 0.06) appears closer to the experimental range, its overall reliability is undermined by the large deviations observed in the modulus predictions.

It should be noted that both DPMD and ClayFF predict moduli that are generally higher than experimental values. This discrepancy can be attributed to several factors.⁵⁸ First, MD simulations are typically based on idealized, homogeneous, and defect-free structural models, whereas real materials contain pores, microcracks, heterogeneous water distributions, and impurities, all of which significantly reduce macroscopic mechanical performance. Second, the preparation methods, curing conditions, and testing protocols of experimental samples strongly influence the measured moduli; for example, water content, curing temperature, and loading rate often lead to lower experimental values. Third, approximations made during the parameterization of forcefields may result in a slight overestimation of bond strength and framework rigidity, thereby yielding higher simulated values. It is worth noting that the density and Poisson ratio were not significantly affected by similar factors. This is mainly because the two properties differ remarkably from the modulus in their physical essence: density, as a fundamental physical property that characterizes the mass per unit volume of a material, has its simulated value dependent only on the spatial arrangement density and atomic mass of atoms, and is less sensitive to microscopic defects (*e.g.*, pores) in the model. Even when an idealized pore-free structure is adopted for simulation, the calculated density can still maintain good consistency with the apparent density of real materials through the direct calculation of system mass and volume. The Poisson ratio reflects the ratio of lateral deformation to longitudinal deformation of a material under stress, and its magnitude is mainly determined by the anisotropy of interatomic bonding. Its susceptibility to macroscopic defects such as pores and microcracks is much lower than that of the



modulus and other mechanical parameters that depend on the overall structural stiffness. Therefore, the aforementioned factors that cause the overestimation of simulated modulus values are not applicable to all performance parameters, which reflects the characteristic differences between different physical quantities in simulation-experiment comparisons.

A comparison between DPMD and ClayFF clearly shows that ClayFF imposes an excessively rigid description of interatomic interactions, leading to predicted moduli that are unrealistically high and physically unreasonable, as well as an overestimated Poisson ratio. In contrast, DPMD, by employing a DPMD model that more accurately fits the energy surface and atomic forces, significantly improves the description of both structural and mechanical properties while maintaining high computational efficiency. Notably, the moduli predicted by DPMD follow the experimental trends, and its dimensionless indicators such as the Poisson ratio are in close agreement with experimental data, further confirming its reliability in complex inorganic polymer systems. Therefore, DPMD not only outperforms the traditional empirical forcefield ClayFF in predicting mechanical properties but also achieves results that are overall closer to experimental observations. From a structural viewpoint, the relatively higher stiffness of N-A-S-H can be attributed to its stronger Na⁺-framework interaction, which promotes a denser packing state, a more constrained local coordination environment, and more stable bridging configurations within the aluminosilicate network. These features collectively enhance resistance to deformation under applied stress. This demonstrates that DPMD provides a more reliable and efficient tool for investigating the structure–property relationships of geopolymer systems, while also revealing the inherent shortcomings of traditional forcefields in modeling complex multicomponent materials and provides a solid theoretical foundation for understanding the structure–property relationships of N-A-S-H materials.

Afterwards, to further validate the generalization capability of the DPMD model, the mechanical properties of the K-A-S-H system are calculated. The DPMD potential is obtained through transfer learning from the N-A-S-H model, in order to examine its applicability in a different cationic environment, as summarized in Table 2.^{59–62} The density of the K-A-S-H system predicted by DPMD is $2.33 \pm 0.03 \text{ g cm}^{-3}$, which lies within the experimental range of $2.23 \pm 0.11 \text{ g cm}^{-3}$, whereas the value obtained from ClayFF ($2.60 \pm 0.07 \text{ g cm}^{-3}$) is noticeably overestimated. The bulk modulus, Young's modulus, and shear modulus measured by DPMD are $29.32 \pm 2.47 \text{ GPa}$, 31.80 ± 2.70

GPa, and $12.05 \pm 1.09 \text{ GPa}$, respectively, which are notably below the respective values from ClayFF ($51.90 \pm 6.07 \text{ GPa}$, $61.20 \pm 5.31 \text{ GPa}$, and $20.51 \pm 2.55 \text{ GPa}$) and much closer to the experimentally reported range.^{59–62} It should be noted that the moduli predicted by DPMD are still slightly higher than experimental values, primarily because the simulated systems are typically idealized, dense, and defect-free, whereas real materials contain pores, microcracks, and heterogeneous water distributions that reduce macroscopic mechanical performance. In addition, experimental conditions such as curing methods and testing protocols can also influence the measured moduli. DPMD demonstrates high accuracy and physical reliability in the K-A-S-H system, thereby confirming the robustness and transferability of the proposed transfer learning strategy for modeling complex geopolymer systems. Compared with N-A-S-H, the lower stiffness of K-A-S-H can be structurally interpreted by its weaker cation-framework coupling, lower packing compactness, and larger free volume. The larger K⁺ ion induces a more weakly connected and more deformable network, which reduces resistance to elastic deformation and results in lower bulk, Young's, and shear moduli. In addition, the mechanical properties of both N-A-S-H and K-A-S-H at elevated temperatures (500–1000 K) are further evaluated, and the results reveal a clear temperature-dependent evolution governed by the competition between intermediate-temperature structural rearrangement and high-temperature thermal disorder; detailed data and analysis are provided in the SI.

3.4. Discussion

In this study, a DPMD model for the N-A-S-H system is successfully constructed within the DP-GEN framework. Building on this, the model is further extended to the K-A-S-H system to examine its generalization capability in a different cationic environment. Through systematic comparisons with AIMD, experimental data, and the conventional forcefield ClayFF, the model is comprehensively validated with respect to energies, forces, structural features, and mechanical properties. Beyond validation, the present results also reveal a clear structure–dynamics–mechanics relationship governed by the charge-balancing cation chemistry. In N-A-S-H, the smaller Na⁺ ion with higher charge density interacts more strongly with framework oxygen atoms, producing a more localized coordination environment, a denser network, and stronger topological constraint. In K-A-S-H, the larger and more weakly bound K⁺ ion leads to broader cation-oxygen coordination, enhanced medium-range structural variability, greater framework flexibility, and lower resistance to deformation. Therefore, the cation effect is not limited to ionic diffusion itself, but propagates across multiple scales to regulate local ordering, network rigidity, ion transport behavior, and ultimately mechanical response. The findings indicate that the DPMD model not only reproduces peak positions, coordination environments, and local structural characteristics with high fidelity to AIMD, but also provides mechanical property predictions that are significantly closer to experimental values than those obtained from ClayFF. Nevertheless, small but observable deviations between

Table 2 The mechanical properties of K-A-S-H obtained from DPMD, ClayFF, and the experiment, respectively

Properties	DPMD	ClayFF	Experiment ^{59–62}
Density (g cm^{-3})	2.33 ± 0.03	2.60 ± 0.07	2.23 ± 0.11
Bulk modulus (GPa)	29.32 ± 2.47	51.90 ± 6.07	27.32 ± 6.58
Young's modulus (GPa)	31.80 ± 2.70	61.20 ± 5.31	26.8 ± 2.2
Shear modulus (GPa)	12.05 ± 1.09	20.51 ± 2.55	10.34 ± 1.26
Poisson ratio	0.32 ± 0.02	0.33 ± 0.03	0.31 ± 0.04



DPMD and AIMD remain in RDF, MSD, and q -order statistics. These deviations can be understood from several aspects. First, although the training dataset spans 300–1000 K and includes diverse local environments, rare or highly distorted configurations may still be underrepresented. Second, the DPMD model remains an approximate neural-network representation of a complex many-body potential energy surface, so a finite fitting error is unavoidable. Third, the effect of thermal fluctuations becomes stronger at elevated temperatures, which amplifies instantaneous differences between DPMD and AIMD, especially in dynamical observables such as MSD and in the broadening of RDF and q -order distributions. Finally, because amorphous geopolymer systems are intrinsically structurally heterogeneous, small local differences in bond lengths, bond angles, and coordination motifs can accumulate into measurable differences in the corresponding statistical descriptors. By incorporating only a limited amount of K-A-S-H training data, the transferred DPMD model achieved accurate predictions of the RDF, MSD, structural order parameter, and mechanical properties. Compared with ClayFF, the DPMD results are in much better agreement with experimental observations, both in terms of peak positions and modulus values. This clearly demonstrates that the transfer learning strategy is capable of substantially lowering the computational cost of developing new models while maintaining high predictive accuracy. The success of this approach highlights the feasibility of applying the DP-GEN framework to cross-system modeling of complex aluminosilicate hydrate gel geopolymers. To further evaluate the applicability of the developed potential beyond the representative composition discussed in the main text, additional validations were performed for systems with different Si/Al ratios, water contents, and alkaline environments, and the corresponding results are provided in the SI. These additional tests show that the developed DPMD model consistently captures composition- and environment-dependent variations in both local structure and tensile response, further supporting its transferability and value for composition-guided geopolymer design.

It should be noted that although the DPMD predictions are overall closer to experimental data, the calculated moduli remain slightly higher than experimental values. This discrepancy primarily arises from the differences between simulated and real materials: MD simulations are typically performed on idealized, dense, and defect-free structures, whereas real geopolymers contain pores, microcracks, heterogeneous water distributions, and impurities, all of which reduce macroscopic mechanical performance. In addition, experimental conditions such as curing methods, temperature, humidity, and loading protocols can significantly influence the measured mechanical properties, often leading to lower values. Despite these differences, the DPMD results still capture the experimental trends and magnitudes with good fidelity, underscoring the reliability of the model in describing realistic chemical environments and mechanical responses.

From a methodological perspective, the successful application of transfer learning is of particular significance. Its significance lies not only in reducing the cost of extending

a pretrained model to a related system, but also in demonstrating a coherent route for constructing system-specific MLP for geopolymer compositions of actual scientific and engineering interest. Compared with conventional forcefield development, where parameter sets are typically fixed and system extension often depends on separate reparameterization or empirical adjustment, the present strategy enables a more flexible and internally consistent description of chemically related amorphous geopolymer gels within the same first-principles-based framework. This is particularly valuable for geopolymer research, where cation chemistry, composition, hydration state, and temperature are all closely coupled to the phenomena under investigation. Traditionally, each new system requires extensive AIMD sampling and model training, which is computationally demanding. The present results show that a DPMD model trained on N-A-S-H can be efficiently adapted to K-A-S-H with only a limited quantity of additional data, while retaining high accuracy. This not only lowers the computational burden but also reveals the scalability and flexibility of deep potential methods in multicomponent systems. Going forward, this approach can be expanded to encompass a wider spectrum of geopolymer compositions, enabling unified modeling across diverse systems that are critical for practical applications. In this broader context, the present work is important not only because it reports accurate models for two representative systems, but because it establishes a transferable route for constructing machine-learning potentials around concrete materials problems, including composition optimization, cation substitution, and temperature-dependent performance prediction in waste-derived, low-carbon geopolymer binders. Specifically, it can be expanded to geopolymers with different alkali cations, like Li^+ and Cs^+ , which are commonly encountered in industrial alkali activators. For these systems, the key lies in leveraging the structural similarity of the aluminosilicate framework while supplementing targeted DFT data that capture the unique coordination behavior and electrostatic interactions of specific cations, especially for cations with distinct ionic radii or charge densities that may induce significant framework relaxation. Furthermore, the approach can be extended to geopolymers with variable Si/Al ratios, which exert a direct impact on the degree of polymerization and mechanical properties of the framework. For such extensions, the training dataset should cover the structural diversity arising from different Si/Al ratios, including variations in tetrahedral connectivity and charge-balancing cation distribution, to ensure that the model generalizes across stoichiometric gradients.

When extending the applicability of this approach to other geopolymers, several critical considerations must be prioritized. Firstly, dataset representativeness is essential; the supplementary DFT data for the target system should cover the key thermodynamic states and structural configurations that are specific to that geopolymer. This ensures that the transfer-learned model captures the unique features of the target system without over-reliance on the parent N-A-S-H model. Secondly, transfer learning parameter optimization is crucial, adjusting the quantity of fixed layers within the neural architecture and



the size of the supplementary dataset based on the similarity between the parent and target systems. For systems closely related to N-A-S-H, fewer frozen layers and a smaller supplementary dataset may suffice; for more distinct systems, increasing the number of trainable layers and expanding the supplementary dataset can improve adaptation. Thirdly, capturing multi-component interactions is necessary for complex geopolymers: the model must be trained to account for non-bonded interactions and chemical bonds involving impurity atoms or additives, which the original model may be unable to capture comprehensively trained on pure N-A-S-H.

Furthermore, as the model continues to be refined, it can be applied to more complex and harsh service environments. For example, under high temperature, humidity, or chemically aggressive conditions, geopolymers undergo significant microstructural evolution, including local crystallization, framework rearrangement, and pore structure changes. These processes are directly linked to long-term durability and service performance. DPMD-based MD simulations can provide atomistic insights into these mechanisms, offering a theoretical basis for understanding degradation behavior under extreme conditions. Such studies are of great importance for applications of geopolymers in nuclear waste immobilization, refractory materials, and long-duration infrastructure. In addition, the high accuracy and efficiency of DPMD models provide a foundation for future multiscale simulations. By coupling atomistic simulations with mesoscale or continuum models, it will be possible to achieve cross-scale investigations from the atomic to the engineering level, thereby enabling a deeper and more integrated understanding of the structure–property correlations of geopolymers. This approach not only helps to reveal fundamental mechanisms but also provides guidance for material design and optimization. With advances in computational power and data availability, it is foreseeable that microstructural simulations with near-DFT accuracy can be realized across the entire geopolymer family, supporting future research on structure–property relationships, durability prediction, and the development of new formulations.

4. Conclusion

In this work, a DPMD model for the N-A-S-H system is developed *via* the DP-GEN framework. Systematic comparisons with AIMD, experimental data, and ClayFF confirm its high accuracy and robustness in terms of energies, forces, structure, and mechanical properties. The model faithfully captures the local structural features and mechanical responses of N-A-S-H while being far more efficient than AIMD for large-scale, extended-timescale simulations. Building on this, a transfer learning scheme extends the model to the K-A-S-H system. With limited K-A-S-H data, the transferred model aligns well with experimental results (including structural characterization, RDFs, MSDs, q -order, and mechanical properties) and significantly outperforms ClayFF. These findings validate transfer learning for cross-system modeling and highlight the strong generalization of deep potential methods in complex inorganic polymers, confirming their reliability in capturing realistic chemical environments and mechanical behaviors. The study also demonstrates the DPMD model's validity in both N-A-S-H and

K-A-S-H systems, as well as transfer learning's great potential in geopolymer research. More importantly, this study demonstrates that MLP can be constructed in a system-specific manner for chemically complex and highly disordered geopolymer gels, providing a flexible alternative to conventional fixed-parameter forcefields. To the best of our knowledge, this work presents the first machine-learning interatomic potentials specifically developed for both N-A-S-H and K-A-S-H systems, and establishes a practical route for extending such models across chemically related geopolymer compositions through a consistent transfer-learning strategy. Future work will extend the model to harsh conditions (*e.g.*, humidity and chemical attack) to study degradation and microstructural evolution (*e.g.*, local crystallization and framework rearrangement). The approach can also be generalized to more geopolymer systems to enable unified modeling across different compositions and ratios. This capability is particularly important for waste-derived, low-carbon binder systems, where composition-dependent structural and mechanical performance must be understood in a coherent framework rather than through disconnected parameter sets. With model refinement and improved computational power, DPMD simulations are expected to achieve near-DFT accuracy across geopolymers, providing a robust theoretical basis for elucidating structure–property relationships, predicting durability, and guiding the design of novel geopolymers.

Author contributions

Yu Li: writing – original draft, methodology, software, investigation, formal analysis, visualization, conceptualization. Jia-ao Hou: writing – original draft, methodology, investigation, formal analysis, visualization, conceptualization. Yuqi Feng: methodology, investigation, visualization, conceptualization. Danyang Zhao: methodology, investigation, formal analysis, conceptualization. Cheuk Lun Chow: writing – review and editing, funding acquisition. Denvid Lau: writing – review and editing, supervision, project administration, funding acquisition, conceptualization.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Supplementary information (SI): details on plane-wave cutoff energy convergence tests, training loss and RMSE evolution of the deep potential models (including transfer learning for K-A-S-H), and comprehensive structural (RDF, MSD, q -order parameters) and mechanical property analyses of N-A-S-H and K-A-S-H geopolymers under varying temperatures, Si/Al ratios, water contents, and alkaline environments. See DOI: <https://doi.org/10.1039/d6ta01407k>.



Acknowledgements

The work described in this paper was fully supported by a grant from the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China (Project No. CityU11213022).

References

- P. Cong and Y. Cheng, Advances in geopolymer materials: A comprehensive review, *J. Traffic Transp. Eng.*, 2021, **8**, 283–314.
- R. M. Novais, R. C. Pullar and J. A. Labrincha, Geopolymer foams: An overview of recent advancements, *Prog. Mater. Sci.*, 2020, **109**, 100621.
- Y. G. Adewuyi, Recent Advances in Fly-Ash-Based Geopolymers: Potential on the Utilization for Sustainable Environmental Remediation, *ACS Omega*, 2021, **6**, 15532–15542.
- J. Stark, Recent advances in the field of cement hydration and microstructure analysis, *Cem. Concr. Res.*, 2011, **41**, 666–678.
- A. M. Elgarahy, A. Maged, M. G. Eloffy, M. Zahran, S. Kharbish, K. Z. Elwakeel and A. Bhatnagar, Geopolymers as sustainable eco-friendly materials: Classification, synthesis routes, and applications in wastewater treatment, *Sep. Purif. Technol.*, 2023, **324**, 124631.
- C. Meral, C. J. Benmore and P. J. M. Monteiro, The study of disorder and nanocrystallinity in C–S–H, supplementary cementitious materials and geopolymers using pair distribution function analysis, *Cem. Concr. Res.*, 2011, **41**, 696–710.
- W. Chen, H. Zhu, Y. Li, F. Liu, Q. Li, Y. Mao and A. Yang, Geopolymers prepared from industrial solid waste: Comprehensive properties and application prospects, *Environ. Res.*, 2025, **278**, 121518.
- R. Abbas, M. A. Abdelzaher, N. Shehata and M. A. Tantawy, Production, characterization and performance of green geopolymer modified with industrial by-products, *Sci. Rep.*, 2024, **14**, 5104.
- K. Półocki, M. Pawlak, J. Drzeżdżon, B. Gawdzik and D. Jacewicz, Clean production of geopolymers as an opportunity for sustainable development of the construction industry, *Sci. Total Environ.*, 2024, **928**, 172579.
- X. Q. Wang, C. L. Chow and D. Lau, Multiscale perspectives for advancing sustainability in fiber reinforced ultra-high performance concrete, *npj Mater. Sustain.*, 2024, **2**, 13.
- H. Hao, S. Li, C. L. Chow and D. Lau, Enhancing compressive strength in cementitious composites via plasma-treated PET aggregates: Insights into interface mechanics, *Cem. Concr. Compos.*, 2024, **149**, 105529.
- M. R. Sadat, S. Bringuier, A. Asaduzzaman, K. Muralidharan and L. Zhang, A molecular dynamics study of the role of molecular water on the structure and mechanics of amorphous geopolymer binders, *J. Chem. Phys.*, 2016, **145**, 134706.
- D. Hou, Y. Zhang, T. Yang, J. Zhang, H. Pei, J. Zhang, J. Jiang and T. Li, Molecular structure, dynamics, and mechanical behavior of sodium aluminosilicate hydrate (NASH) gel at elevated temperature: a molecular dynamics study, *Phys. Chem. Chem. Phys.*, 2018, **20**, 20695–20711.
- H. Hao, C. L. Chow and D. Lau, Effect of heat flux on combustion of different wood species, *Fuel*, 2020, **278**, 118325.
- H. Hao, C. L. Chow and D. Lau, Carbon monoxide release mechanism in cellulose combustion using reactive forcefield, *Fuel*, 2020, **269**, 117422.
- H. L. Hao, R. Y. Qin, C. L. Chow and D. Lau, A multiscale model for wood combustion, *Comput.-Aided Civ. Infrastruct. Eng.*, 2024, **39**, 2903–2916.
- D. Zhao, X. Q. Wang, L.-h. Tam, C. L. Chow and D. Lau, Tailored twisted CNT bundle with improved inter-tube slipping performances, *Thin-Walled Struct.*, 2024, **196**, 111536.
- X. Q. Wang, P. Chen, C. L. Chow and D. Lau, Artificial-intelligence-led revolution of construction materials: From molecules to Industry 4.0, *Matter*, 2023, **6**, 1831–1859.
- Y. Zhang, T. Li, D. Hou, J. Zhang and J. Jiang, Insights on magnesium and sulfate ions' adsorption on the surface of sodium alumino-silicate hydrate (NASH) gel: a molecular dynamics study, *Phys. Chem. Chem. Phys.*, 2018, **20**, 18297–18310.
- Y. Zhao, J. Du, X. Qiao, X. Cao, C. Zhang, G. Xu, Y. Liu, S. Peng and G. Han, Ionic self-diffusion of Na₂O–Al₂O₃–SiO₂ glasses from molecular dynamics simulations, *J. Non-Cryst. Solids*, 2020, **527**, 119734.
- D. Hou, J. Zhang, W. Pan, Y. Zhang and Z. Zhang, Nanoscale mechanism of ions immobilized by the geopolymer: A molecular dynamics study, *J. Nucl. Mater.*, 2020, **528**, 151841.
- L.-Y. Xu, Y. Alrefaei, Y.-S. Wang and J.-G. Dai, Recent advances in molecular dynamics simulation of the N-A-S-H geopolymer system: modeling, structural analysis, and dynamics, *Constr. Build. Mater.*, 2021, **276**, 122196.
- Y. Feng, H. Hao, C. L. Chow and D. Lau, Exploring reaction mechanisms and kinetics of cellulose combustion via ReaxFF molecular dynamics simulations, *Chem. Eng. J.*, 2024, **488**, 151023.
- T. Wang, X. He, M. Li, Y. Li, R. Bi, Y. Wang, C. Cheng, X. Shen, J. Meng, H. Zhang, H. Liu, Z. Wang, S. Li, B. Shao and T.-Y. Liu, Ab initio characterization of protein molecular dynamics with AI2BMD, *Nature*, 2024, **635**, 1019–1027.
- Y. Feng, S. Mekhilef, D. Hui, C. L. Chow and D. Lau, Machine learning-assisted wood materials: Applications and future prospects, *Extreme Mech. Lett.*, 2024, **71**, 102209.
- W. Li, C. Xiong, Y. Zhou, W. Chen, Y. Zheng, W. Lin and J. Xing, Insights on the mechanical properties and failure mechanisms of calcium silicate hydrates based on deep-learning potential molecular dynamics, *Cem. Concr. Res.*, 2024, **186**, 107690.
- S. Choung, W. Park, J. Moon and J. W. Han, Rise of machine learning potentials in heterogeneous catalysis:



- Developments, applications, and prospects, *Chem. Eng. J.*, 2024, **494**, 152757.
- 28 Y. Zhang, H. Wang, W. Chen, J. Zeng, L. Zhang, H. Wang and W. E, DP-GEN: A concurrent learning platform for the generation of reliable deep learning based potential energy models, *Comput. Phys. Commun.*, 2020, **253**, 107206.
- 29 J. Zeng, D. Zhang, A. Peng, X. Zhang, S. He, Y. Wang, X. Liu, H. Bi, Y. Li, C. Cai, C. Zhang, Y. Du, J.-X. Zhu, P. Mo, Z. Huang, Q. Zeng, S. Shi, X. Qin, Z. Yu, C. Luo, Y. Ding, Y.-P. Liu, R. Shi, Z. Wang, S. L. Bore, J. Chang, Z. Deng, Z. Ding, S. Han, W. Jiang, G. Ke, Z. Liu, D. Lu, K. Muraoka, H. Oliaei, A. K. Singh, H. Que, W. Xu, Z. Xu, Y.-B. Zhuang, J. Dai, T. J. Giese, W. Jia, B. Xu, D. M. York, L. Zhang and H. Wang, DeepMD-kit v3: A Multiple-Backend Framework for Machine Learning Potentials, *J. Chem. Theory Comput.*, 2025, **21**, 4375–4385.
- 30 X. Luo, X. Tian, J. Wu, X. Yang, Z. Liu, Z. Jiao and H. Peng, Molecular simulations of the initial stage's induction and formation process of N-A-S-H Gel based on NaOH-activated Metakaolin, *J. Non-Cryst. Solids*, 2024, **626**, 122804.
- 31 M. R. Sadat, K. Muralidharan and L. Zhang, Reactive molecular dynamics simulation of the mechanical behavior of sodium aluminosilicate geopolymer and calcium silicate hydrate composites, *Comput. Mater. Sci.*, 2018, **150**, 500–509.
- 32 K. Zhu, Performance Comparisons of NequIP and DPMD Machine Learning Interatomic Potentials for Tobermorites, *Comput. Mater. Sci.*, 2024, **244**, 113212.
- 33 Y. Zhou, H. Zheng, W. Li, T. Ma and C. Miao, A deep learning potential applied in tobermorite phases and extended to calcium silicate hydrates, *Cem. Concr. Res.*, 2022, **152**, 106685.
- 34 E. Samaniego, C. Anitescu, S. Goswami, V. M. Nguyen-Thanh, H. Guo, K. Hamdia, X. Zhuang and T. Rabczuk, An energy approach to the solution of partial differential equations in computational mechanics *via* machine learning: Concepts, implementation and applications, *Comput. Methods Appl. Mech. Eng.*, 2020, **362**, 112790.
- 35 J. Huang, L. Zhang, H. Wang, J. Zhao, J. Cheng and W. E, Deep potential generation scheme and simulation protocol for the Li₁₀GeP₂S₁₂-type superionic conductors, *J. Chem. Phys.*, 2021, **154**, 094703.
- 36 D. K. Sharma, M. Chatterjee, G. Kaur and S. Vavilala, 3 – Deep learning applications for disease diagnosis, in *Deep Learning for Medical Applications with Unique Data*, ed. D. Gupta, U. Kose, A. Khanna and V. E. Balas, Academic Press 2022, pp. 31–51.
- 37 G. Cai, D. Zhang, J.-a. Hou, D. Lau, R. Qin, W. Wang, W. Zhang, C. Wu and L.-h. Tam, Machine learning prediction models for investigating vibration properties of epoxy resin under moisture conditions, *Int. J. Non-Linear Mech.*, 2024, **166**, 104857.
- 38 R. Wu, X. Q. Wang, D. Zhao, J.-a. Hou, C. Wu, D. Lau and L.-h. Tam, Degradation of fiber/matrix interface under various environmental and loading conditions: Insights from molecular simulations, *Constr. Build. Mater.*, 2023, **390**, 131101.
- 39 C. Wu, J.-a. Hou, H. Liu, J. Yang, D. Lau and L.-h. Tam, Understanding moisture effect on nonlinear vibrations of epoxy thin film *via* a multiscale simulation, *J. Sound Vib.*, 2023, **553**, 117649.
- 40 Y. Li and D. Lau, Advances in shape memory polymers and their composites: From theoretical modeling and MD simulations to additive manufacturing, *Giant*, 2024, **18**, 100277.
- 41 M. Bu, Q. Yang, P. Wang, B. Dong, D. Hou and Y. Wang, Dissolution behaviors and mechanisms of metakaolin in acidic activators, *Cem. Concr. Res.*, 2024, **178**, 107442.
- 42 E. K. Goharshadi, A review on the radial distribution function: Insights into molecular structure, intermolecular interactions, and thermodynamic properties, *J. Mol. Liq.*, 2025, **433**, 127900.
- 43 M.-F. Kai and J.-G. Dai, Understanding geopolymer binder-aggregate interfacial characteristics at molecular level, *Cem. Concr. Res.*, 2021, **149**, 106582.
- 44 E. Duque-Redondo, F. Basquiroto de Souza, G. Geng and H. Manzano, A critical review and perspectives on atomistic models of non-crystalline cementitious materials, *Cem. Concr. Res.*, 2026, **199**, 108067.
- 45 B. Dianat, F. Tavanti, A. Padovani, L. Larcher and A. Calzolari, BELLO: A post-processing tool for the local-order analysis of disordered systems, *Comput. Mater. Sci.*, 2022, **209**, 111381.
- 46 F. Tavanti and A. Calzolari, Concurring effect of doping and composition on the thermodynamic properties of amorphous GexSe_{1-x} alloys, *Acta Mater.*, 2024, **266**, 119676.
- 47 B. Dianat, P. La Torraca, A. Manfredi, G. Cassone, C. Vacchi, M. Sebastiani and F. Pancaldi, Classification of pulmonary sounds through deep learning for the diagnosis of interstitial lung diseases secondary to connective tissue diseases, *Comput. Biol. Med.*, 2023, **160**, 106928.
- 48 T. Williamson, T. Zhu, J. Han, G. Sant, O. B. Isgor, M. C. G. Juenger and L. Katz, Effect of temperature on N-A-S-(H) and zeolite composition, solubility, and structure, *Cem. Concr. Res.*, 2023, **172**, 107213.
- 49 Y. Liu, X. Hu, Y. Du, B. Nematollahi and C. Shi, A review on high-temperature resistance of geopolymer concrete, *J. Build. Eng.*, 2024, **98**, 111241.
- 50 H. Y. Zhang, V. Kodur, B. Wu, L. Cao and F. Wang, Thermal behavior and mechanical properties of geopolymer mortar after exposure to elevated temperatures, *Constr. Build. Mater.*, 2016, **109**, 17–24.
- 51 P. Li, Y. Li, W. Ou and X. Ran, Conceptual design and properties of ultra-high residual strength geopolymer after 1000 °C thermal exposure, *Mater. Lett.*, 2024, **373**, 137153.
- 52 H. Kucukgoncu and A. Özbayrak, Microstructural Analysis of Low-Calcium Fly Ash-Based Geopolymer Concrete with Different Ratios of Activator and Binder Under High Temperatures, *Arabian J. Sci. Eng.*, 2025, **50**, 8197–8223.
- 53 B. Ren, J. Wang, Z. Zhou, P. Du and X. Zhang, Regulation of the composition of metakaolin-based geopolymer: Effect of zeolite crystal seeds, *Case Stud. Constr. Mater.*, 2023, **19**, e02421.



- 54 M. Criado, W. Aperador and I. Sobrados, Microstructural and Mechanical Properties of Alkali Activated Colombian Raw Materials, *Materials*, 2016, **9**, 158.
- 55 W. Li, Y. Wang, C. Yu, Z. He, C. Zuo and Y. Yu, Nano-scale study on molecular structure, thermal stability, and mechanical properties of geopolymer, *J. Korean Ceram. Soc.*, 2023, **60**, 413–423.
- 56 G. A. Lyngdoh, S. Nayak, N. M. A. Krishnan and S. Das, Fracture toughness of fly ash-based geopolymer gels: Evaluations using nanoindentation experiment and molecular dynamics simulation, *Constr. Build. Mater.*, 2020, **262**, 120797.
- 57 J. Němeček, V. Šmilauer and L. Kopecký, Nanoindentation characteristics of alkali-activated aluminosilicate materials, *Cem. Concr. Compos.*, 2011, **33**, 163–170.
- 58 L.-h. Tam, R. Wu, J.-a. Hou, C. Wu, Basics of Molecular Dynamics Simulation Methods, in *Molecular Simulation Investigations of Property Degradation in CFRP Composite*, ed. L.-H. Tam, R. Wu, J.-A. Hou and C. Wu, Springer Nature Singapore, Singapore, 2024, pp. 35–51.
- 59 R. Cai, T. Wu, C. Fu and H. Ye, Thermal degradation of potassium-activated ternary slag-fly ash-silica fume binders, *Constr. Build. Mater.*, 2022, **320**, 126304.
- 60 J. Tailby and K. J. D. MacKenzie, Structure and mechanical properties of aluminosilicate geopolymer composites with Portland cement and its constituent minerals, *Cem. Concr. Res.*, 2010, **40**, 787–794.
- 61 H. Gao, R. Yu, J. Liu, W. Zhang, M. Li, Y. Jia, Y. Ye, S. Zhong, C. Zhuang, H. Zhu, Q. Su, B.-T. Huang, H. Wu, J. Sun and D. Hou, Study on the molecular structure and mechanical properties of potassium ion uptake in calcium silicate hydrate, *Constr. Build. Mater.*, 2025, **481**, 141651.
- 62 N. Sedira, J. Castro-Gomes and M. Magrinho, Red clay brick and tungsten mining waste-based alkali-activated binder: Microstructural and mechanical properties, *Constr. Build. Mater.*, 2018, **190**, 1034–1048.

