



A molecular dataset for the shear deformation of thermoplastic structural materials

Gao Yuzhao¹, Li Chaochao¹, Zhang Rui², Ma Yuanyuan^{3,*}, Lu Kuan^{2,*}

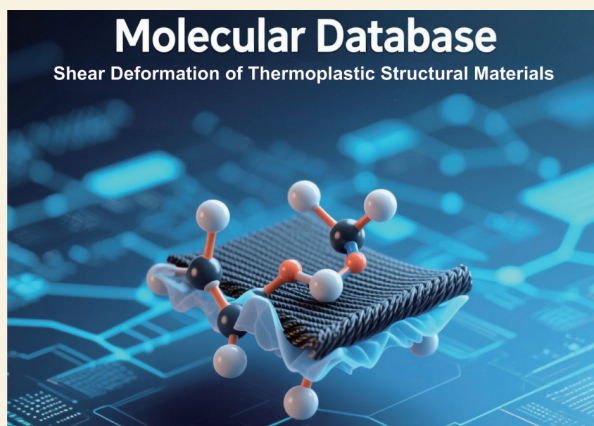
(1. School of Statistics, Shanxi University of Finance and Economics, Taiyuan 030006, China;

2. Institute of Coal Chemistry, Chinese Academy of Sciences, State Key Laboratory of Coal Conversion, Taiyuan 030001, China;

3. School of Materials Science and Engineering, North University of China, Taiyuan 030051, China)

Abstract: The first molecular dynamics (MD) simulation dataset is reported for the interfacial shear behavior of carbon fiber/thermoplastic composites (CFRTPs), aimed at overcoming the critical interfacial problem that limits their high-end applications such as aerospace and new energy vehicles. The study features 2 key advances. First, we use the newly developed CHONSi-2024 reactive force field (ReaxFF), which provides high-precision parameters specifically for CFRTP interfacial systems; Second, the atomic models are constructed based on experimental characterization data and rigorously validated across multiple parameters, including shear modulus, yield behavior, stress-strain curves, and fracture morphology, ensuring quantitative agreement with experimental results. This dataset provides a complete record of the simulations, encompassing atomic trajectories, local structural changes, interfacial stress-strain responses, and system thermodynamic behaviors. These data offer direct atomic-scale insights into the interfacial strengthening mechanisms. The generated trajectories are compatible with mainstream software for visualization and analysis. Moreover, the dataset constitutes a high-quality resource for developing machine learning force fields, building structure-property relationships, and enabling the predictive modeling and high-throughput screening of composite interfaces. This dataset is anticipated to advance the fundamental understanding of composite interfaces and facilitate the rational design of high-performance CFRTPs.

Key words: Carbon Fiber; Polyamide; Composites; ReaxFF; Interface



1 Introduction

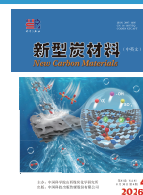
Carbon fiber-reinforced thermoplastic composites (CFRTPs) are a new generation of lightweight structural materials with significant application potential in aerospace, new-energy vehicles, and high-value-added equipment. Their ultra-high specific strength and stiffness, coupled with excellent fatigue resistance, make them critical for achieving lightweight design and sustainable industrial development. Compared with conventional thermosetting composites, CFRTPs offer distinct advantages including excellent recyclability, high molding efficiency and outstanding impact toughness, which are highly compatible with the global green manufacturing develop-

ment strategy. However, the large-scale industrial application of CFRTPs still significantly lags behind that of their thermosetting counterparts. The critical bottleneck is the high melt viscosity of thermoplastic matrices, which leads to insufficient interfacial wetting, a high propensity for interfacial defect formation, and failure to achieve the theoretically predicted interfacial shear strength (IFSS), therefore severely restricting the full realization of their comprehensive properties. Accordingly, interfacial regulation has

Received: January 12, 2026

Revised: April 21, 2026

Accepted: April 21, 2026



emerged as a core research direction in this field. A robust interfacial structure ensures efficient stress transfer between fibers and matrix, enabling the full exertion of the reinforcing effect of carbon fibers. Therefore, an in-depth elucidation of interfacial interaction mechanisms is crucial for the performance optimization of high-performance CFRTPs.

Traditional interfacial research primarily relies on experimental approaches including characterizing fiber surface morphologies and chemical states by microscopic and spectroscopic techniques (e.g., scanning electron microscopy (SEM), atomic force microscopy (AFM), X-ray photoelectron spectroscopy (XPS)), and assessing interfacial properties through macroscopic mechanical tests (e.g., micro-droplet debonding tests for IFSS quantification). While these methods can capture macroscopic interfacial phenomena, they fail to dynamically resolve the interfacial interaction process at the atomic scale. In particular, there is still a lack of in-depth quantitative insights into the dynamic formation, spatial distribution of interfacial hydrogen bond networks, and their contribution to stress transfer efficiency.

In this work, the thermoplastic matrices investigated are types of polymers: polyamides and polyurethane. Specifically, a polyamide (PA3) with a polymerization degree of 3 serves as a sizing agent and another polyamide (PA6) with a polymerization degree of 6 serves the matrix, while a polyurethane (PU3) with a polymerization degree of 3 is exclusively employed as a sizing agent. To address the aforementioned technical and scientific challenges, we constructed a shear deformation dataset of carbon fiber/polyamide (polyurethane) composites based strictly on experimental characterization data, with 2 major innovative breakthroughs. (1) Conventional IFSS data are mainly derived from macroscopic experimental measurements, which cannot uncover the dynamic details of the interfacial failure process. In this dataset, the entire interfacial shear debonding process is dynamically reproduced at the atomic scale by molecular dynamics (MD) simulations in this field. (2) Traditional interfacial studies often vaguely ascribe the strengthening

mechanism to “chemical bonding” or “mechanical interlocking”, lacking quantitative characterization of key microscopic parameters. Based on the atomic structures in this dataset, the density of the interfacial hydrogen bond network, as a core microscopic descriptor, can be directly and precisely quantified. This high-fidelity simulation dataset lays a solid theoretical foundation for the efficient research and development of high-performance CFRTPs, and promotes a paradigm shift in composite material development from empirical process optimization to mechanism-driven rational design.

Unlike widely used datasets such as the Materials Project, which mainly focuses on the bulk properties of crystalline materials, this dataset is the first specialized database for atomic-scale simulation of interfacial shear behavior in carbon fiber/polymer composites. Its unique advantage lies in providing full-chain data covering atomic-scale structural evolution to macroscopic mechanical responses, offering an exclusive data resource for in-depth elucidation of interfacial failure mechanisms and the realization of machine learning-assisted rational design of composite interfaces.

2 Simulation process and results

2.1 Data generation method and process

The generation of simulation data can be systematically divided into 2 key stages: model construction and simulation execution. First, in the model construction stage, instead of using idealized smooth surfaces, the study constructed carbon fiber surface models with controllable roughness based on surface roughness data provided by AFM using mathematical functions, accurately replicating the surface topography differences between dry-jet wet spinning carbon fibers (DS-CFs) (Fig. S1) and wet spinning carbon fibers (WS-CFs) (Fig. S2). Subsequently, based on surface chemical information obtained from XPS analysis (Table S1), specific amounts of oxygen-containing functional groups (e.g., $-\text{COOH}$ and $-\text{OH}$) were quantitatively grafted onto the rough surface model, ensuring the accuracy of the surface chemical state in

the model (Table 1). On this basis, sizing agent molecular chains (PA3 and PU3) were grafted onto the functionalized fiber surface, and finally, a matrix region containing 200 PA6 molecular chains was assembled, constructing a highly realistic three-layer composite interface model of fiber/sizing agent/matrix (Fig.1). In other words, the dataset includes 2 independent sizing agent systems (PA3 and PU3) with PA6 as the unified matrix.

Second, simulations were performed using the CHONSi-2024 force field^[1] and the LAMMPS software^[2] specifically developed for carbon fiber/matrix composites. For single-molecule models, structural optimization was first performed, and the resulting structures were used to construct subsequent composite fiber models. For composite models, structural optimization was initially performed on the fiber composite model, followed by relaxation to a density of 1.12 g/cm³ and a temperature of 298 K in the Isothermal-Isobaric Ensemble (NPT ensemble). Then, further structural relaxation for 200 ps was carried out in the Canonical Ensemble (NVT ensemble). This relaxation process ensured sufficient contact and interaction between the PA/PU molecular chains and the fiber surface. The relaxed model was used for subsequent shear simulation. The critical shear deformation simulation was carried out at 600 K with approximately 51 000 atoms under a shear rate of 10 m/s and an integration timestep of 0.25 fs using the Verlet integration algorithm. The complete dynamic process of interfacial debonding was simulated by fixing the carbon fiber atoms and applying a constant radial shear rate to the PA6 matrix until complete separation of fiber and matrix was achieved (about 1.4 ns). Structural visualization can be implemented using OVITO software^[3].

Notably, the elevated temperature (600 K) was

adopted to accelerate the kinetic process of interfacial debonding, while the core microscopic mechanism of interfacial shear deformation is consistent with that at room temperature, which has been verified by the good agreement between the simulated fracture morphology and experimental observations. The shear rate in conventional mechanical tests is typically on the order of 10^{-6} – 10^{-3} m·s⁻¹, whereas the high shear rate in the simulations was set to accelerate the interfacial debonding process and enable the observation of the complete failure process within a feasible simulation time. This high shear rate may lead to a slight overestimation of the yield stress, but the evolution trend of the stress-strain curve and the core interfacial failure mechanisms are highly consistent with experimental results. The system size (~ 51 000 atoms) is limited by current computational resources, and the relatively small system may not fully reproduce the macroscopic scale effect of actual composites. However, the atomic-scale interfacial interaction mechanisms (e.g., hydrogen bond network evolution and polymer chain adsorption) are not affected by the system size, which has also been validated by the consistency between simulated and experimental fracture morphologies. The carbon fiber atoms were fixed to simplify the simulations and focus on the deformation and failure behavior of the polymer matrix and the interface, which is a common practice in MD simulations of fiber/polymer composites. This treatment may neglect the slight deformation of the carbon fiber surface, but carbon fibers have an ultra-high modulus and their deformation is negligible during the actual interfacial shear process, thus exerting a minimal impact on the simulation results.

2.2 Objectives and advantages of the dataset

The core objective of the MD simulations in this work is to construct atomic models that can accur-

Table 1 Four surface groove models representing different surface roughness levels and their surface functional group distribution

Surface feature	DS-CFs		WS-CFs	
	Surface smooth, minor grooves exist (Fig. S1)		Surface has grooves of varying depth (Fig. S2)	
Groove model	Minor groove(A1)	Small groove(A2)	Medium groove(A4)	Large groove(A8)
Surface functional groups	—C—C	421 bridge-site C atoms		
	—C—OH	213 top-site OH groups		
	—COOH	5 top-site COOH groups		

ately reproduce the realistic interfacial structure of composites, and to elucidate the microscopic mechanism governing interfacial performance at the atomic scale by simulating the interfacial shear process. The molecular dataset for shear deformation of carbon fiber/polyamide (polyurethane) composites established in this work realizes the precise correlation between macroscopic IFSS and key microscopic structural descriptors by ReaxFF MD simulations, thus constructing complete data spanning “atomic configuration-molecular interaction-macroscopic mechanical properties”. This dataset overcomes the limitations of traditional studies that rely solely on empirical correlations and provides a fundamental basis for establishing predictive models for the rational design of high-performance CFRTPs. The dataset has profound academic and application value. On the one hand, it can provide direct quantitative guidance for optimizing the heat treatment process of CFRTPs to achieve the controllable enhancement of interfacial properties. On the other hand, the structured and standardized atomic trajectory data and structure-property mapping data serve as high-value resources for training machine learning potentials, developing specialized large language models for materials science, and building intelligent knowledge question-and-answer systems. This dataset also provides a solid datum infrastructure for accelerating the development of novel composites and constructing digital teaching cases for computational materials science.

2.3 Data sample description

The dataset contains 8 subfolders in total: Box-A1-100PA3-200PA6 (containing 14 334 .data files), Box-A1-100PU3-200PA6 (14 109 .data files), Box-A2-100PA3-200PA6 (17 588 .data files), Box-A2-100PU3-200PA6 (15 799 .data files), Box-A4-100PA3-200PA6 (13 338 .data files), Box-A4-100PU3-200PA6 (15 224 .data files), Box-A8-100PA3-200PA6 (11 849 .data files), and Box-A8-

100PU3-200PA6 (14 862 .data files). Here, A1, A2, A4 and A8 represent the 4 model types: minor groove, small groove, medium groove and large groove, respectively. 100PA3 and 100PU3 represent 100 molecules of PA3 and PU3, respectively, representing the sizing agent layer. 200PA6 represents 200 molecules of PA6, representing the matrix layer. Each folder contains three types of text files: .data files, .SS files, and .log files. The specific file contents are detailed in Table 2.

2.4 Data quality

The reliability of the data are ensured by a multi-level verification and control framework employed in the molecular dynamics simulations. First, a validated high-precision force field^[1] provided the fundamental description of interatomic interactions. Rigorous thermodynamic equilibrium and convergence controls were implemented during the simulation: a Langevin thermostat maintained constant temperature, while the two-dimensional (2D) temperature distribution within the moving region was calculated to verify the correct velocity profile. Additionally, a sufficiently long equilibrium period (100 ps) was used, and statistical noise was further reduced by time-averaging — sampling every 2 steps, accumulating 100 samples for averaging, and reporting results every 200 steps. Ultimately, the stability of the simulation was supported by the trend of total energy evolution during shear (Fig. 2) for the Box-A4-100PA3-200PA6 model.

Second, theoretical results were validated against experimental data through multi-scale cross-verification. The force field parameters used in the simulations were calibrated against experimental elastic constants, lattice constants and thermodynamic properties. The calculated shear modulus, yield behavior and simulated stress-strain curves were consistent in trend and magnitude with experimental curves. Quantitatively, the simulated IFSS values for the DS-CF and

Table 2 Storage contents of the molecular dataset for shear deformation of carbon fiber/polyamide (polyurethane) composites

Data type	Data content	Data format	Software
Structural information	Evolution of reactive interface over time, including periodic boundary conditions, atomic coordinates, etc.	.data	LAMMPS
Shear force statistical information	Contains stress and strain information.	.SS	
Thermodynamic statistical data	Contains timestep, temperature, energy, etc.	.log	

WS-CF models were 101.34 and 69.97 MPa, respectively. These values were slightly higher than the optimized experimental values of 80.10 MPa for DS-CFs and 61.30 MPa for WS-CFs (Fig. 1a)^[4]. This discrepancy was logically attributed to the idealized conditions inherent to molecular dynamics simulations, which model a nearly perfect interface devoid of real-world defects. Consequently, the simulated IFSS values can be interpreted as representing the theoretical upper limit of interfacial bonding strength achievable for these composites. The advantage of the constructed groove models is that they are established through mathematical functions based on experimental surface roughness data, enabling the quantitative control of surface roughness parameters (amplitude/wavelength) and eliminating the interference of random factors on real fiber surfaces. This simplified model can accurately capture the key physical effects of surface roughness on interfacial shear performance (e.g., mechanical interlocking), and enable the quantitative analysis of the relationship between surface roughness and IFSS. Meanwhile, the regular groove

model ensures high reproducibility of the simulation results. The limitations of the model are as follows: the actual surface of sized carbon fibers presents irregular and random micro-nano topography, and the distribution of the sizing agent is heterogeneous due to the influence of surface energy and wettability. The simplified regular groove model cannot fully reproduce the randomness of the actual surface topography and the non-uniform distribution of the sizing agent, which may lead to slight differences between the simulated local stress distribution and the actual situation. In addition, the model does not consider the micro-defects (e.g., microcracks) that may exist in the actual sizing layer, which is another reason for the slightly higher simulated IFSS than the experimental value.

Moreover, the final simulated structures were fully consistent with experimental observations. Upon complete interfacial debonding, more PA6 matrix residue remains on the DS-CF surface than on the WS-CF surface. This indicates stronger interfacial bonding between DS-CFs and PA6 compared to WS-CFs, consistent with the observed interfacial morpho-

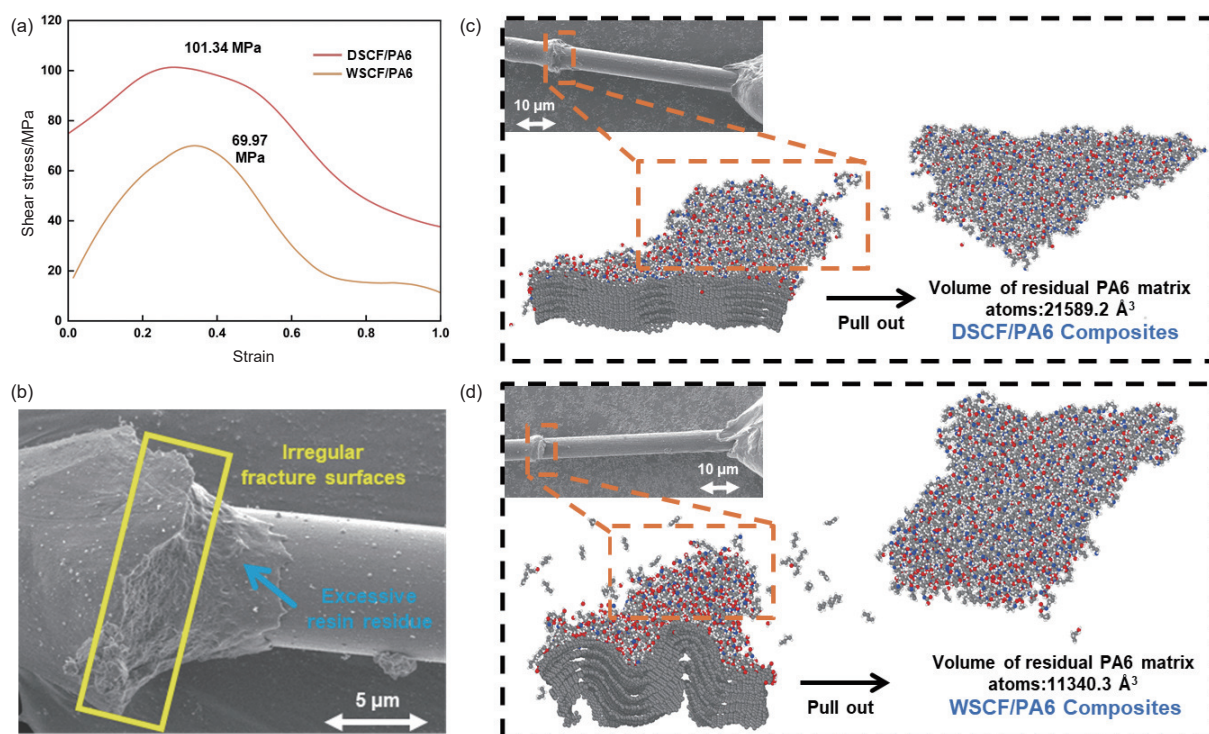


Fig. 1 (a) Simulated stress-strain curves for the DS-CFs (101.34 MPa) and WS-CFs (69.97 MPa) composite. (b) Fracture surface morphology. Schematic of shear fracture consistent with experimental phenomena for (c) DS-CF composite and (d) WS-CF composite. Gray: carbon atoms, red: oxygen atoms, white: hydrogen atoms, blue: nitrogen atoms

logy (Fig. 1b). In addition, the hydrogen bond density of the DS-CF structure (11.06 bonds/nm², Fig. 3) is higher than that of the WS-CF structure (9.61 bonds/nm²), which is the core difference contributing to the superior interfacial performance of DS-CFs over WS-CFs. This multi-level, self-consistent validation system ensures the reliability and physical authenticity of the theoretical data, laying a solid foundation for mechanistic analysis.

2.5 Data usage methods and suggestions

The atomic structure trajectories and other output data generated by these simulations can be used for multi-dimensional analysis, supporting the investigation of material mechanical behavior and underlying microscopic mechanisms. Specific application scenarios and suggestions are as follows:

First, the visualization and fundamental analysis of atomic trajectories constitute to the foundation for subsequent work. It is recommended to use the OVITO software to load the trajectory files and to perform structural analysis utilizing its built-in tools. For visualization purposes, simply opening the structure files allows for the intuitive observation of the dynamic evolution of shear deformation. For quantitative analysis, tools such as “Coordination number analysis” or “Radial distribution function” can be computed to quantitatively characterize the evolution of local atomic order before and after shear. This can be combined with the “Slice” function to observe structural changes within targeted regions of interest.

Second, it is necessary to conduct a correlation

analysis between macroscopic mechanical properties and microscopic atomic processes. The macroscopic mechanical response can be obtained directly from the IFSS data. It is suggested to plot shear stress-strain curves to analyze properties such as yield strength (Fig. 1a). The key step is to precisely correlate characteristic points on the macroscopic curve (e.g., yield point, stress drop) with microscopic events observed in OVITO (e.g., polymer chain scission, changes in hydrogen bond density) at corresponding time points. Furthermore, tools like “Atomic Strain” can be applied to dynamically moving atom groups to quantitatively map the evolution of strain localization zones. Combined with defect analysis to track the generation of vacancies or interstitial atoms, this approach can reveal the atomistic chemical processes underlying plastic deformation.

Finally, to establish predictable structure-property relationships, it is recommended to construct machine learning models based on atomic trajectories and characteristic parameters. For example, we converted the .data files (stored in LAMMPS dump format) into a standardized extended XYZ format. This is implemented by a custom Python script, executing the command: `python dump2exyz-multi.py --log Box-A4-100PA3-200PA6_xShear_600K_0.0001.log --dump dump --output Box-A4-100PA3-200PA6_xShear_600K_0.0001.xyz --interval 100` processes the dump/timestep_*.data files, converting atomic coordinates and thermodynamic data (linked to log-derived energies) into a unified trajectory file. We used

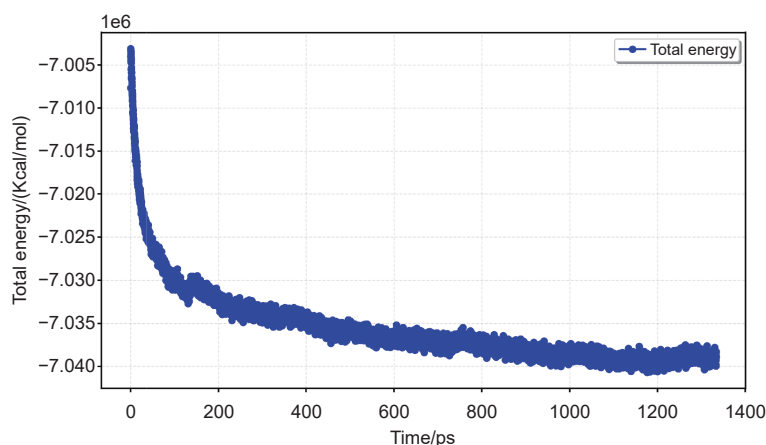


Fig. 2 Variation trend of total energy with shear simulation for the Box-A4-100PA3-200PA6 model

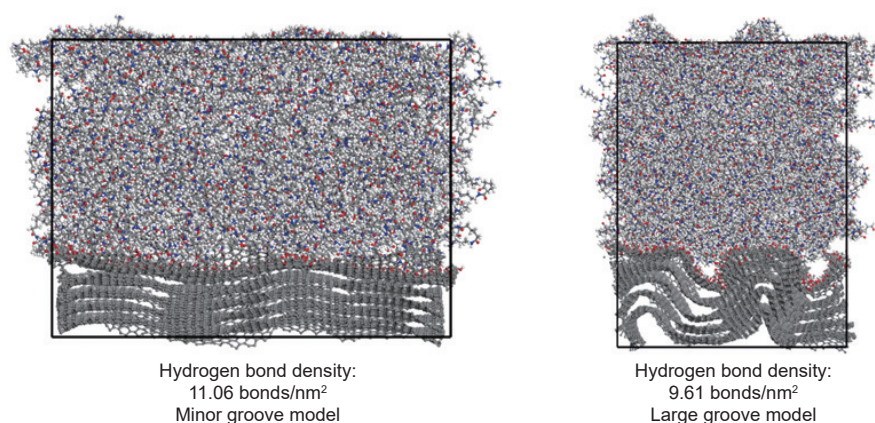


Fig. 3 Atomic schematic of different groove models with different hydrogen bond densities. Note: The minor groove model is denoted as A1, and the large groove model is denoted as A8. Hydrogen bond densities: 11.06 bonds/nm² for the minor groove model and 9.61 bonds/nm² for the large groove model.

Gray: carbon atoms, red: oxygen atoms, white: hydrogen atoms, blue: nitrogen atoms

the smooth overlap of atomic positions (SOAP) descriptor to extract the structural features of the atomic configurations in the dataset, and then performed the dimensionality reduction analysis by PCA and t-SNE. The generated XYZ file (e.g., Box-A4-100PA3-200PA6_xShear_600K_0.0001.xyz, including 133 structures and associated energy data by interval 100) was then analyzed using a dedicated script `soap.py`, which focuses on structural and energetic characterization through dimensionality reduction algorithms. Specifically, principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) were employed to visualize the distribution of atomic configurations and their associated energies. PCA highlights differences between systems through linear dimensionality reduction, capturing global variance in structural features, while t-SNE emphasizes local correlations among configurations (such as the clustering of reaction intermediates in t-SNE space) via nonlinear mapping. The PCA similarity plot (Fig. 4a) visually demonstrates structural similarity through the spatial distribution of points. For t-SNE (Fig. 4b), the core advantage lies in preserving local similarity, directly correlating with structural stability. The strain can be divided into 3 main stages: (1) the 0-0.6 stage represents the high-stress phase of the structure, with correspondingly higher energy distribution, particularly in the 0.2-0.4 range, indicating the maximum stress distribution; (2) the 0.6-1.1 stage represents the gradual reduction in stress, corresponding to the progressive debonding process of the com-

posite material; (3) when the strain exceeds 1.1, the composite interface is almost completely fractured. The results demonstrate that the PCA/t-SNE analysis constructed based on the dataset can quickly and accurately identify the three key stages of interfacial shear deformation. Other local structural features of each atom at a specific timestep, such as bond orientation order parameters, atomic volume and local stress, can also serve as input features. The phase to which an atom belongs (crystalline/defect/amorphous) or its future motion can be used as labels for training classification or regression models. Feature engineering is crucial: descriptors sensitive to shear deformation should be selected, and data should be standardized. Models trained in this manner can predict plastic behavior from an initial atomic configuration or identify “warning” atoms critical for shear band nucleation, thereby deepening the understanding of the initiation mechanisms of plastic instability.

3 Conclusions

We have constructed and comprehensively described a multi-scale correlated molecular simulation dataset focused on the interfacial shear behavior of carbon fiber/polyamide (polyurethane) composites through systematic ReaxFF MD simulations. The dataset comprehensively captures full-chain data spanning atomic trajectories, local structural evolution, interfacial stress-strain responses and system thermodynamic parameters. The atomic trajectory files are stored in a standard format, which can be dir-

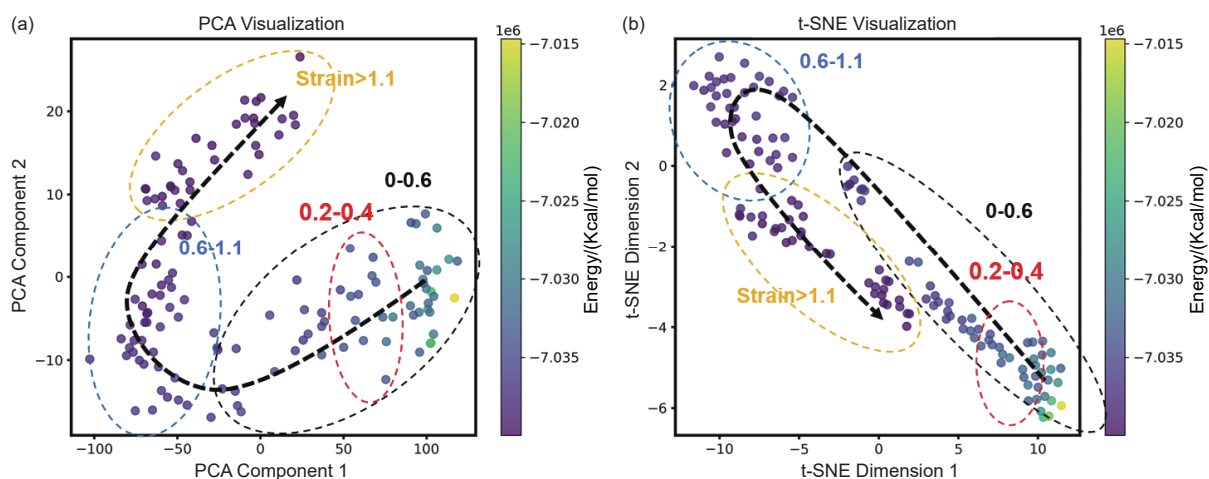


Fig. 4 Cluster analysis of 133 generated structural and energy data by (a) PCA and (b) t-SNE methods from the Box-A4-100PA3-200PA6 model. The black arrow represents the direction of time

ectly used for dynamic visualization and quantitative analysis with mainstream simulation and visualization software, facilitating the intuitive interpretation of key processes including interfacial structure evolution and polymer chain rearrangement. More importantly, the structured feature data extracted from the dataset (including local order parameters, atomic stress/strain and coordination environments) act as ideal input for training machine learning force fields and establishing quantitative structure-property relationship models. This provides high-quality training resources for developing intelligent computational models that can predict the mechanical properties of composite interfaces, identify weak atomic configurations, and realize high-throughput screening of interfacial structures. This dataset is expected to accelerate the rational design of novel high-performance CFRTPs, to promote the development and refinement of interfacial shear deformation theory, and to provide a key data foundation for the application of materials informatics in the field of high-performance composites.

References

- [1] Lu K, Ma Y, Lu C, et al. Exploring the atomistic-scale formation and evolution of rings in PAN-based carbon fibers using a reparameterized ReaxFF approach[J]. *The Journal of Physical Chemistry C*, 2025, 129(30): 13670-13683.
- [2] Plimpton S. Fast parallel algorithms for short-range molecular dynamics[J]. *Journal of computational physics*, 1995, 117(1): 1-19.
- [3] Stukowski A. Visualization and analysis of atomistic simulation data with OVITO—the Open Visualization Tool[J]. *Modelling and simulation in materials science and engineering*, 2010, 18(1):

015012.

- [4] Zhang R, Zhi J, Cao L, et al. Unraveling the interfacial hydrogen bond network mechanism in carbon fiber/polyamide composites: across-scale investigation of carbon fiber with different roughness for interface optimization[J]. *Composites Science and Technology*, 2025: 111453.