Exploring Thermoset Fracture with A Quantum Chemically Accurate Model of Bond Scission

Zheng Yu and Nicholas E. Jackson*

Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, USA

E-mail: jacksonn@illinois.edu

Abstract

A molecular understanding of thermoset fracture is crucial for enhancing performance and durability across applications. However, achieving accurate atomistic modeling of thermoset fracture remains computationally prohibitive due to the high cost associated with quantum mechanical methods for describing bond breaking. In this work, we introduce an active learning (AL) framework for our recently developed machine-learning based adaptable bond topology (MLABT) model that uses datasets generated via density functional theory (DFT) calculations that are both minimalistic and informative. Employing MLABT integrated with AL and DFT, we explore fracture behavior in highly crosslinked thermosets, assessing the variations in fracture behavior induced by system temperature, temperature fluctuations, strain rate, cooling rate, and degree of crosslinking. Notably, we discover that while fracture is minimally affected by temperature, it is strongly influenced by strain rate. Furthermore, while the structural disparities introduced by different network annealing rates influence the elastic properties, they are inconsequential for thermoset fracture. In contrast, network topology emerges as the dominant determinant of fracture, influencing both the ultimate strain and stress. Particularly, MLABT with AL-DFT achieving near quantum-chemical accuracy in bond breaking still leads to ductile failures, emphasizing the necessity of modeling polymer networks at larger length scales for bridging the gap between experiment and simulation. Nevertheless, the integration of MLABT with the AL framework paves the way for efficient and DFT-accurate modeling of thermoset fracture, providing an affordable and accurate approach for calculating polymer network fracture across chemical space.

Introduction

Thermosets, characterized by the presence of irreversible polymer crosslinks and enhanced mechanical properties, are foundational to numerous technological applications ranging from automotive components and aerospace structures to medical devices and protective coatings.[1–3] The robust nature of thermosets along with their adaptability have ushered in new horizons for material innovations.[4,5] Central to maximizing the potential of thermosets in these domains is an in-depth understanding of their fracture behaviors.[6–10] Recent advancements in computational tools and experimental techniques have provided insights into the fracture of thermosets.[11–18] However, the atomic-scale processes governing these behaviors remain less explored. Delving into this atomic realm promises not only enhanced material predictability but also the prospect of tailored design.[19,20] Nevertheless, this pursuit presents formidable challenges: accurate atomic-level modeling of thermoset fracture necessitates substantial computational resources, especially when elucidating intricate bond breakage phenomena with quantum mechanical (QM) methods.[21] As the demand for higher performance materials grows, overcoming these challenges and obtaining a more comprehensive understanding of thermoset fracture at the molecular scale is imperative.

In our previous study, we introduced the Machine Learning based Adaptable Bonding Topology (MLABT) framework, an approach tailored for atomistic simulations of thermosets under large deformation.[22] MLABT circumvents limitations of classical molecular dynamics (MD) simulations by itegrating a machine learning (ML) algorithm for detection and execution of bond-breaking events (with near QM accuracy) with any underlying classical force-field. Compared to existing methods combining MD and QM,[21] MLABT exhibits an approximately two orders of magnitude improved computational efficiency, coupled with heightened sensitivity to rare bond-breaking events at low strains. This blend of speed and accuracy created by augmenting classical force-fields accurate at low strain with QM-quality bond-breaking detection renders MLABT a robust and chemically general tool for probing strain hardening and material failure dynamics in polymer networks.

While the development of MLABT is promising for modeling thermoset fracture, it is not without its challenges. First and foremost, generating a sufficient volume of training data using QM calculations is resource-intensive. Moreover, provided the rare nature of bond-breaking events, bond evaluations are primarily confined to a narrow strain-hardening window for efficiency reasons. Second, existing MLABT training data derives from configurations possessing intact cross-linked topologies, rather than those emerging *in situ* during fracture. These constraints, revolving around inefficient and insufficient configurational sampling, curtail the full potential of MLABT in offering a computationally efficient and molecularly detailed modeling paradigm for thermoset fracture. Recognizing these limitations, the field of active learning (AL),[23,24] defined by its ability to iteratively refine and expand training datasets, appears an apt fit for these challenges.[25,26]

In this work we integrate AL into MLABT simulations. This provides two concerted benefits for MLABT models: the ability (i) to systematically navigate the vast configurational space of thermoset fracture and (ii) to employ higher accuracy (i.e. more computationally costly) density functional theory (DFT) calculations that better capture the physics of bond breaking, by virtue of using nearly an order of magnitude less training data than previously. Empowered by this AL-DFT MLABT model, we conduct the first comprehensive exploration of thermoset fracture behaviors at the molecular scale using a DFT-accurate bond-breaking model. Note that although classical material fracture involves crack propagation beyond the molecular scale, we use the more specific term “fracture” instead of “failure”, because breaking of the network into two parts can be clearly observed in the simulations, and it is consistent with pre-existing literature studying similar phenomena.[6,7,9] We scrutinize how the stress-strain behaviors as well as bond breakages are modulated by experimental factors including temperature, strain rate, cooling rate, and the degree of crosslinking. The results from our MLABT analysis reveal that strain rates have a more pronounced effect on bond breakages than temperature, suggesting that the time-temperature superposition principle does not hold for thermoset plasticity. Importantly, network topology appears to be more influential than the stability of the glassy structure in determining fracture behavior. Furthermore, we demonstrate that thermoset stress-strain behavior exhibits small variance under thermal fluctuations, indicating a certain degree of degeneracy in network fracture. Considering that MLABT leverages the DFT accuracy for bond breaking in fracture modeling, yet doesn't capture the brittle failure seen in experiments, we conjecture that embracing larger spatiotemporal scales in modeling will be essential for better alignment with experiments.

Methods

MLABT simulation

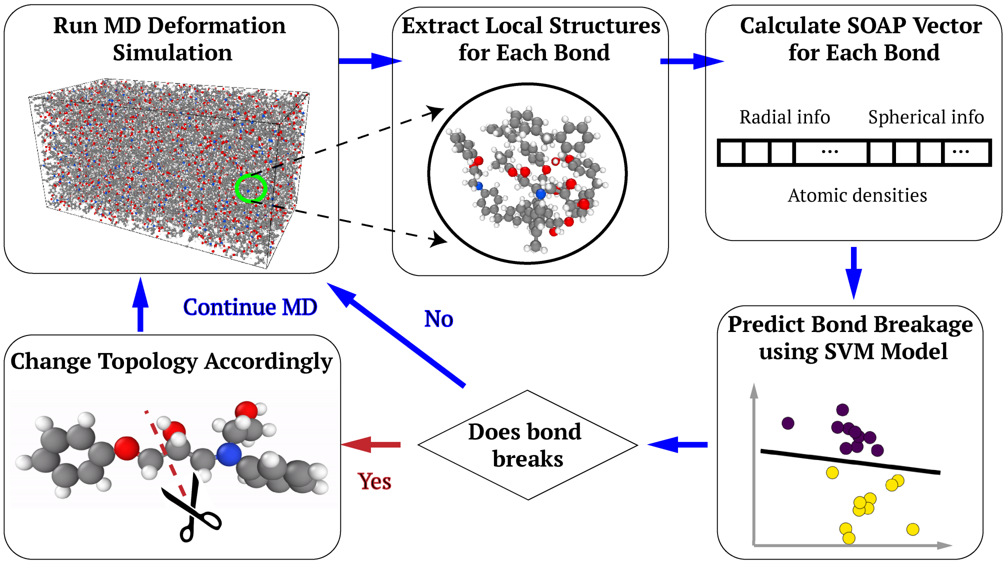

Figure 1: Schematic of the Machine-Learning-based Adaptable Bonding Topology (MLABT) method. MLABT can efficiently predict and perform bond breaking on-the-fly in MD simulations with near quantum-chemical accuracy.[22]

MLABT is a method incorporated on-the-fly with classical MD (e.g. OPLS, Amber) to accurately describe quantum-chemically accurate bond breaking at dramatically reduced cost, with a focus on the modeling of thermoset deformation and fracture.[22] As illustrated in Fig. 1, MLABT scans all potentially breakable bond types in the classical MD simulation and predicts bond rupture based on the local structures. If a bond breaks, the corresponding topology is automatically modified and MD continues until the next bond breaks. We develop MLABT in an archetypal epoxy polymer network, diglycidyl ether of bisphenol A (DGEBA) cured by methylene dianiline (MDA), but the approach is chemically generalizable.

We employ similar simulation parameters as described in our previous work.[22] Specifically, a cubic box containing 432 DGEBA and 216 MDA molecules (27,432 atoms in total) is utilized with periodic boundary conditions in three dimensions. Bonding topologies of networks are generated dynamically by simulating curing reactions in MD, resulting in degrees of crosslinking ranging from 77% to 98%. Structures are melted at 800 K for 200 ps and then quenched to 300 K with a constant annealing rate ranging from 0.1 K/ps to 100 K/ps. The obtained glassy structures are then used as initial conditions for MLABT deformation simulations. Only uniaxial deformations are considered in this work. During deformations, the simulation box is deformed uniaxially (e.g., along the x axis) every 0.025 ps at a strain rate of 1×10^9/s and the atomic coordinates are remapped accordingly. The two transverse directions are allowed to relax under P=1 atm to avoid the accumulation of artificial stress. We apply the Optimized Potentials for Liquid Simulations All Atoms (OPLS-AA) force-field with the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) in all MD simulations.[27,28] Simulated glass transition temperature, density, and elastic properties are all in good agreement with experiments and previous simulations.[11,29–31]

Central to MLABT is the ML model tasked with predicting bond breakages by analyzing the bond's instantaneous surroundings. To characterize the local structure, we employ the Smooth Overlap of Atomic Positions (SOAP) descriptor, representing a Gaussian smeared local atomic density based on spherical harmonics and radial basis functions.[32–34] We apply the support vector machine (SVM) with the radial basis function kernel as the classifier.[35] More details can be found in the reference.[22] Considering the additional cost of ML prediction that requires the computation of the SOAP vectors for all the relevant atoms, we perform the scanning of bond breaking every 0.001 true strain. In our testing, as long as the evaluation frequency is greater than one check every $\Delta\varepsilon$ =0.01, no evident difference is observed in the resultant deformation behavior, as illustrated in Fig. S3 of the Supporting Information. In addition, due to the instability of the structures with broken bonds simulated with OPLS-AA, a timestep of 0.25 fs is utilized in MLABT simulations. Tuning force fields or adding hydrogens to broken bonds could solve this instability issue, but is not performed in the present study. All the simulations conducted in this work are carried out on the Bridges-2 cluster, which is provided by the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS).[36]

Iterative MLABT-based active learning

To improve the generalization ability of MLABT across the entire fracture process, the ML model in this work undergoes iterative refinement, enriched progressively by the incorporation of AL. AL is a data-driven methodology primarily aimed at optimizing the process of data labeling and model training.[37] Distinct from traditional ML frameworks, where a model is trained on a pre-labeled dataset, AL centers on the model actively selecting the most informative data points from a vast unlabeled pool of data.[38,39] Here, we harness the power of AL to elucidate the fracture behaviors of thermosets, minimizing costs of QM computations while maximizing predictive generalization ability at diverse deformation conditions.

The overall AL workflow is illustrated in Fig. 2. We start sampling highly strained configurations in MD simulations and extracting the local structures that potentially contain broken bonds for QM geometry optimization.[22] Data pre-screening requires an artificial threshold based either on bond length or stretching energy, which could limit the applicability of the ML model in early bond breaking prediction. To achieve a high fidelity model, we utilize the more accurate DFT method PBEh-3c that improves upon our previous work using the semi-empirical tight-binding method GFN2-xTB.[22,40] Due to the increased computational cost of PBEh-3c, our computational budget permitted generating a smaller initial training dataset, containing around 5,000 data points.

The initial SVM model is fit using 80% of the data and applied to kickstart the AL campaign.

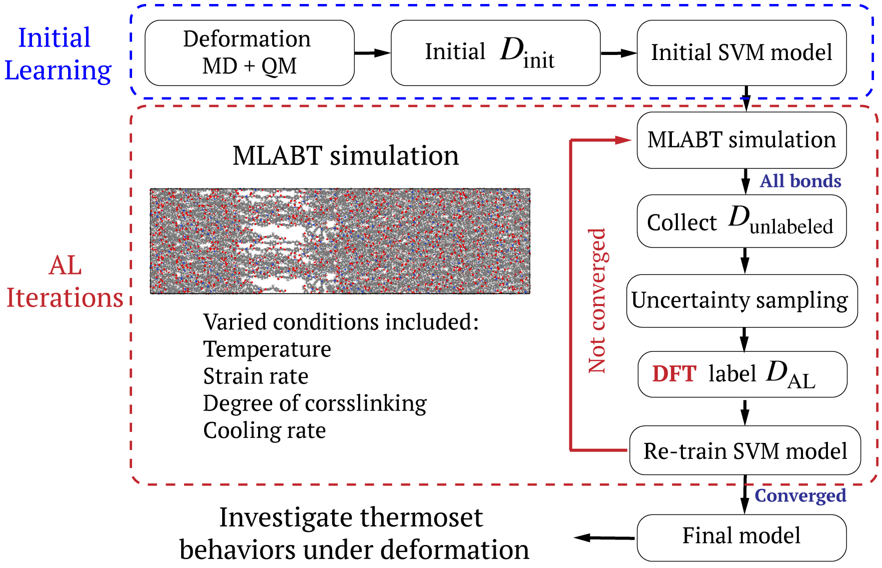

Figure 2: Workflow of the iterative active learning framework for MLABT.

To include representative and diverse local structures during deformation into the model training, we employ an iterative pool-based AL strategy, as illustrated in Fig. 2. In each AL iteration, we collect all inputs (SOAP vectors) of the potentially breakable bonds (around 4,000 bonds per frame) in all frames (around 1,250 frames per trajectory) of a MLABT trajectory, which is generated based on the latest re-trained ML model, as an unlabeled data pool. Note that this step takes no additional computational cost since SOAP vectors of these bonds were already computed while performing MLABT simulations. By doing this, the unlabeled data pool in one iteration already contains comprehensive information of in situ chemical bond local environments within a broad range of strains throughout the thermoset fracture process. To further improve the model's generalization ability, we introduce some variations in conditions of MLABT simulations during the AL iterations. In the second iteration, we include simulations at various temperatures from 100 K to 400 K. In the third iteration, we include simulations starting from initial structures with different bonding topologies. In the fourth iteration, we include simulations with different strain rates and with initial structures in different degrees of crosslinking (78% to 98%). Over all iterations, around 85 million unlabeled data instances are collected cumulatively for AL querying.

Active learning query strategy

To select the most informative data instances (local configurations) from the large pool of unlabeled data for DFT labeling, we use uncertainty sampling, along with the SVM classifier. In the context of SVM, uncertainty sampling can be intuitively understood by examining the decision function for each prediction.[41,42] For a data point, the absolute value of the decision function $|f(x)| = |\sum_i^N \alpha_i \gamma_i \kappa(x_i, x) + b|$ represents its distance to the decision boundary, where $\kappa(x_i, x) = \langle \phi(x_i) \phi(x) \rangle$ is a kernel function and $\alpha_i \gamma_i \phi(x_i)$ forms a weight vector. The smaller this absolute value, the closer the data point is to the decision boundary, which indicates a higher level of uncertainty. Thus, querying data points with the smallest absolute decision functions maximizes the information gain, refining our model with each iteration.

However, an inherent challenge emerges when adopting this approach: as our model and data evolve across iterations, the absolute values of the decision function can shift, rendering them non-comparable across different AL cycles. This poses a problem when trying to maintain a consistent measure of uncertainty across multiple iterations. To circumvent this challenge, we employ Platt scaling—a method wherein a logistic regression model is trained using the decision function's outputs.[43] Through this process, the SVM's raw decision values are transformed into calibrated probabilities, providing a consistent measure of uncertainty irrespective of the active learning iteration. Within our binary classification context $P(1|x) = 1 - P(-1|x)$, the uncertainty associated with each instance x (SOAP vectors) is defined as

$$u(x) = 1 - \max(P(1|x), P(-1|x)) = 0.5 - |P(1|x) - 0.5| \tag{1}$$

Under this formulation, data points with probabilities closer to 0.5 are deemed to have maximum uncertainty, as they lie in regions where the model is most uncertain about its classifications.

Using the outlined query strategy, bonds with maximum uncertainty are identified within each snapshot of the MLABT simulation. To regulate the number of bonds selected in each AL iteration, we apply an uncertainty threshold of 0.05. For every selected bond, its local environment is extracted from the large MD configuration. This isolated environment is then subjected to full optimization via DFT calculations, as in the initial dataset generation.[44] Each batch of labeled data from the AL is partitioned into a training set (comprising 80% of the data) and a testing set (accounting for the remaining 20%). Subsequently, the SVM model is retrained, incorporating both the initial dataset and the cumulative new AL training data. Test data are composed of both the initial data and the cumulative AL test data, as well as unseen data from a new MLABT trajectory (with maximum uncertainty in every snapshot) based on the final model. Detailed results on the model convergence are presented in the subsequent section.

Results

Active learning performance

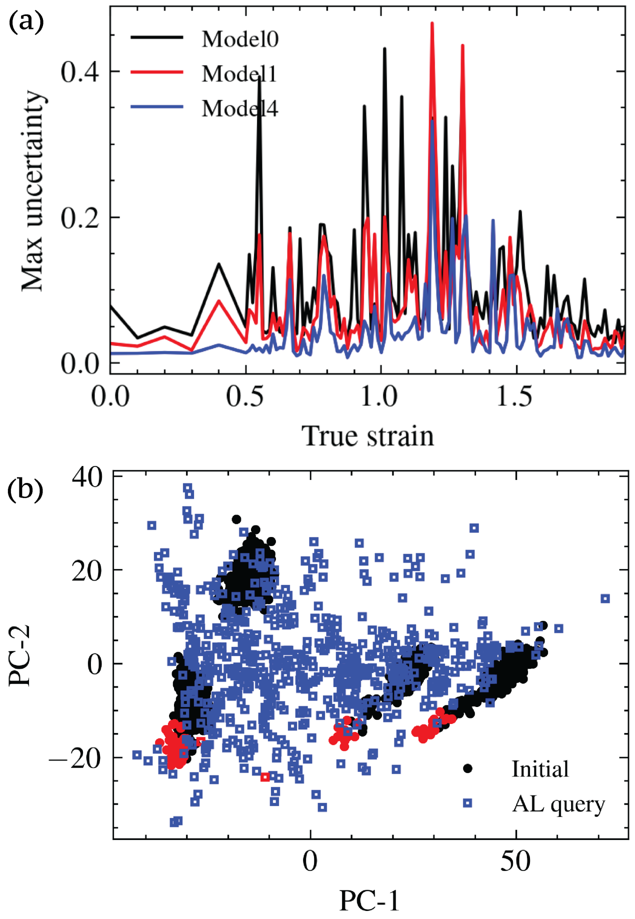

Figure 3: (a) Max uncertainty of bond breaking prediction in MLABT simulations after AL iterations. Uncertainty of bond breaking prediction is computed based on Equation 1, as detailed in the Methods section. ‘Model 0’ denotes the model trained by the initial data, ‘Model 1’ and ‘Model 4’ denotes the updated models after the first and fourth AL iteration, respectively. (b) Initial data and AL data visualized by principle component analysis of the SOAP vectors. The red points represent bonds that are found broken in DFT calculations. The evident difference in data distributions demonstrates that AL explores diverse regions in feature space that are distinct from the initial sampling.

The AL framework in this work is based on the query strategy of uncertainty sampling, as detailed in the Methods section. Since bond breaking events are rare even in material fracture, the majority of bond breaking uncertainties are simply zero, even for the initial ML model with a small dataset, as shown in Fig. S1 of the Supporting Information. However, the maximum uncertainty during the deformation could be high if the bond instance lies closer to the poorly trained decision boundary. As shown in Fig. 3A, the initial model shows high peaks in uncertainty around strains of 0.5-0.6, where the bonds start to break, and around strain of 0.9-1.3, where the bonds break rapidly and the resulting stress reaches a maximum. The strain region in between has relatively lower uncertainty, because it is where the initial data are generated. The uncertainty for larger strains ($>$1.3) decreases, due to weaker interaction between strained bonds in the system that is poorly crosslinked. Nevertheless, the bonds with maximum uncertainty above a threshold of 0.05 in each snapshot are selected, and their local configurations are optimized with DFT to determine bond breaking.

Next, we compare the local environments of the AL selected bonds with the bonds in the initial dataset, which are selected by bond stretching energy from configurations in a narrow range of strain, in the reduced dimensions by principle component analysis (PCA). The linear transformation (coefficients) is constructed based on the SOAP vectors of the bonds in the initial dataset. As shown in Fig. 3B, the local environments in the initial dataset form four separated clusters. The right two clusters are associated with the 'CT-CA' bonds (connecting the sp3 carbon and the aromatic carbon) located on both DGEBA and MDA, and the left two clusters are associated with the 'CT-CT' bonds on DGEBA. Note that only those 'CT-CT' bonds on the DGEBA backbone (the bottom left cluster) are breakable in deformation, while those on DGEBA side chains (the top left cluster) are not. We could remove those from training data, but in this work, they are kept to enhance generalizability. On the other hand, the local environments selected by AL are scattered over the principal component space and distributed densely in regions between the clusters. This result demonstrates the ability of AL to explore the diverse feature space that is unseen in the initial data. As such, including these AL environments in the model training can improve the generalization ability of MLABT for simulating thermoset deformations under diverse conditions.

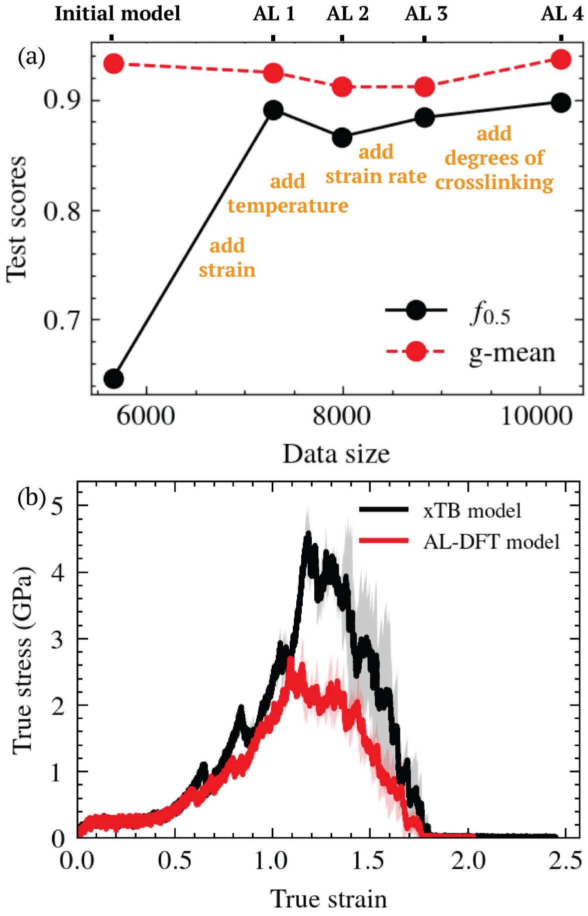

Figure 4: (a) $f_{0.5}$ and geometric mean of the ML model as trained after each AL iteration. (b) Comparison of stress-strain curves simulated by AL-DFT MLABT model and with the xTB MLABT model at 300 K with a strain rate of 10^9/s. The xTB MLABT model overestimates the ultimate stress, compared to the AL-DFT model and previous simulations.[45] The shadow regions represent the standard deviations induced by randomness in initial velocity generation over three independent trajectories (same for subsequent figures).

Figure 4A shows the performance of the ML model on the test set after each AL iteration. The selection of the scoring metrics considering the imbalanced classification was discussed in previous work.[22,46] It can be seen that the initial model exhibits an excellent geometry mean, whereas $f_{0.5}$ is relatively low, indicating a higher rate of false positives on regions outside of the initial training region. Once the model is updated with the AL selected data from the entire strain range using varied deformation and temperatures, $f_{0.5}$ improves immediately while geometric mean remains almost unchanged. With more AL iterations including deformations using various strain rates and configurations with various degrees of crosslinking, the performance of the model remains almost unchanged, whereas the maximum uncertainty decreases evidently (Fig. 3). After the fourth AL iteration, the maximum uncertainty is almost zero before the first bond breaking, and it remains above 0.15 only in a narrow strain range around 1.2. This performance is reflected in the distribution of prediction uncertainty, as shown in Fig. S1B and S1C of the Supporting Information. In addition, the number of bonds in the same deformation trajectory with uncertainty above 0.05 decreases with more AL iterations, but the rate of decrease slows after the second iteration, as shown in Fig. S2A of the Supporting Information. These features all suggest that the model in the AL framework is converging. To confirm the convergence, we apply the models after each iteration in MLABT simulations with identical initial conditions (positions, velocities). The results of the models after the second iterations are very similar, especially at $\varepsilon < 1.5$, as shown in the Fig. S2B of Supporting Information. As such, we end the AL campaign and employ the model after the fourth iteration as the final model (denoted as 'AL-DFT') in this work for further investigation.

MLABT simulations with the new AL-DFT MLABT model provide more accurate results on thermoset fracture than with the previously reported 'xTB' MLABT model. As shown in the stress-strain curves (SSC) in Fig. 4B, although the two models produce similar strains for the fracture initiation (the first bond breaking), the ultimate stress, and the material failure, the ultimate stress of AL-DFT MLABT is roughly one half of that of xTB MLABT, showing improved agreement with previous simulations (1-3 GPa).[45] This reduced ultimate stress is induced by an increase of broken bonds and already appears in the initial model, suggesting that the GFN2-xTB method compared to DFT underestimates the probability of bond breaking. In addition, we find that simple models only using bond length as the breakage criterion, as reported in earlier simulations,[21,47] could delay early bond breaking events and fail to generalize across various thermodynamic or morphological conditions, as detailed in the Supporting Information (Fig. S4). Furthermore, the shadow regions in Fig. 4B shows the standard deviation caused by randomness in velocity initialization, i.e., random seeds in generating initial atom velocities from the Maxwell–Boltzmann distribution. Although the specific broken bonds and locations are different, the overall variance in the SSC during strain hardening is small. This suggests that there is some degeneracy in fracture initiation sites during thermoset deformation that result from velocity initialization, but they weakly affect the overall mechanical properties. The variation of ultimate stress and failure is stronger, probably due to the accumulated differences in bond breaking resulting in evident differences in the broken topology. We note that to our knowledge this study represents the first QM-informed atomistic study to report error bars in stress-strain curves of thermoset fracture, as for traditional approaches such error bars would be too computationally costly to compute.

Bond breaking during deformation

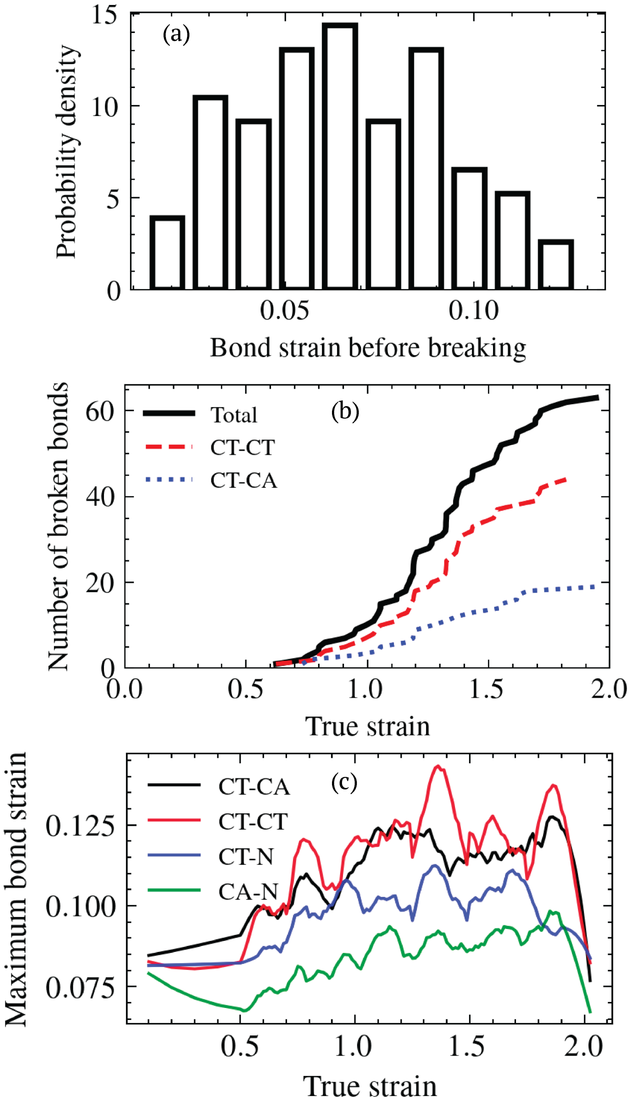

Figure 5: (a) Bond strain distribution immediately before bond breakage during MLABT simulations in the highly crosslinked system (98%) at 300 K with a strain rate of 10^9/s. (b) The numbers of broken bonds in the two dominant broken bond types as a function of strain. (c) The maximum bond strains for various bond types in the system as a function of strain. The bond strain is defined as $(l - l_0)/l_0$, where l is the instantaneous bond length and l_0 is the equilibrium bond length.

The more accurate AL-DFT MLABT model enables investigation of the bond breaking chemistry during thermoset fracture. Figure 5A shows the distribution of bond strain, defined as $(l - l_0)/l_0$, where l is the instantaneous bond length and l_0 is the equilibrium bond length, immediately before bond breaking during MLABT simulations in the highly crosslinked system (98%) at 300 K with a strain rate of 10^9/s. Similar results are observed at other conditions. It is surprising that the bond breakages occur at much smaller bond strains than previously expected.[48] The maximum value around 0.12 is even smaller than the threshold value used in Barr's method for pre-screening local configurations that potentially contain broken bonds,[21] suggesting that the methods based on artificial criteria and QM calculations can delay bond breaking during deformation.

The actual types of broken bonds in MLABT simulations of thermoset fracture are found to align with those revealed in QM calculations. Note that although we observed two types of broken bonds in strained DGEBA+MDA by QM calculations (both DFT and xTB), it does not guarantee these two types are actually broken during thermoset fracture because the local environments with evolving topology could become very different. As shown in Fig. 5B, the numbers of broken bonds decomposed into the two types show a consistent ratio throughout the entire fracture process. This ratio, i.e., approximately 2:1 for 'CT-CT' and 'CT-CA', is in agreement with the observations in the DFT calculations. This suggests that the bond breaking mechanism in epoxy thermosets remains consistent during fracture, independent of global strain. Furthermore, because our ML model is designed for only predicting these two types, we need to evaluate the possibility of bond breaking in other bond types during fracture. Figure 5C illustrates the maximum bond strain for four types of potentially broken bonds in the entire MLABT simulation. The 'CT-CT' and 'CT-CA' bonds indeed exhibit the highest maximum bond strain, with 'CT-CT' frequently being slightly higher than 'CT-CA'. This is consistent with the fact that only these two types were broken and that 'CT-CT' bonds have a higher probability of rupture. The maximum bond strain of 'CT-N' or 'CA-N' are evidently lower and cannot exceed those of the two broken types in the entire fracture range, confirming that they cannot break and negligibly contribute to the ML bond breaking model.

Fracture behaviors by MLABT

Utilizing AL-DFT MLABT simulations, we can efficiently probe the fracture behaviors of polymer networks at the atomic scale, combining the molecular precision and computational efficiency of classical MD with bond-breaking fidelity approaching that of DFT. It is imperative to recognize, however, that both the lengthscale and timescale exert significant influence on the network topology's formation and its dynamical response. We focus on extracting physical insights, specifically examining how the bond breaking events and fracture behaviors of polymer networks are modulated by factors such as temperature, strain rates, cooling rates, and the degree of crosslinking.

Temperature effect

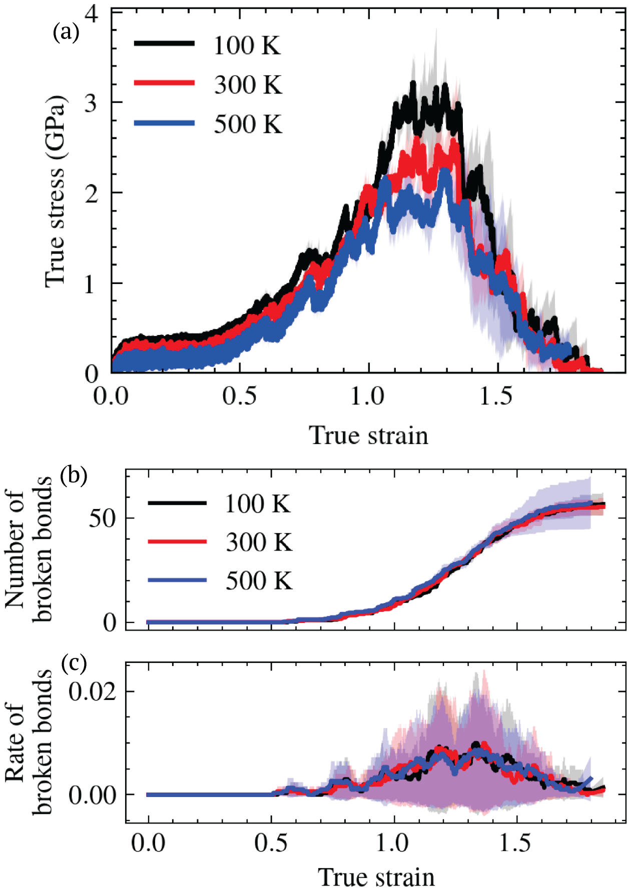

Figure 6: (a) Stress-strain curves of the same initial structure (98% crosslinked) simulated by AL-DFT MLABT at three temperatures with the same strain rate of 10^9/s. (b)-(c) Corresponding accumulated number and rate of the bond breaking events during deformation. The unitless rate in (c) is simply the derivative of the accumulated number in (b), with smoothing and normalization for better visualization. The shadow regions represent the standard deviations induced by randomness in initial velocity generation over three independent trajectories, and the curves represent the averaged results. The same settings are applied in subsequent figures.

First, we assessed the impact of temperature, held constant during deformation, on the fracture behavior of thermosets. Figure 6 displays (a) the SSC, (b) the count of broken bonds, and (c) the rate of bond breakage (as deduced from the slope in (b)) at three distinct temperatures, 100 K, 300 K, and 500 K, in MLABT simulations of a 98% crosslinked system subjected to a strain rate of 10^9/s. These temperatures are all below the glass transition temperature ($T_g \sim 541$ K). In general, bond breaking events appear temperature-insensitive, with temperature only influencing the ultimate stress as a consequence of the temperature effect before yielding. Specifically, a decrease in temperature enhances the elastic modulus and the corresponding yield stress, in agreement with previous experiments and simulations.[49] However, the frequency of bond breakage events remain consistent in the plastic regime, and consequently, the characteristic strains for fracture initiation, peak stress, and ultimate failure also exhibit temperature independence. Only the stresses during strain softening and hardening vary as a result of the effect on the elastic regime, which was also observed in previous MD simulations without considering bond breaking.[50] These findings suggest that bond rupture in amorphous polymer networks may not be characterized as a simple activation reaction. The potential reason may be related to heterogeneous local stresses that arise depending on the global strain and the network topology.[51,52] Additional temperature-dependent behaviors in experiments, such as increased brittleness at lower temperatures,[53] could be attributed to factors like crystallinity, entanglements, or effects occurring over extended spatiotemporal scales.

Moreover, we find that the resultant stress is generally correlated with the bond breakage rate, a trend consistently observed across all conditions evaluated in this study. As the count of bonds on the edge of breaking surges during strain hardening, the cumulative stress also rises until the bond breakage rate peaks, resulting in the ultimate stress. Following this, as the bond breaking rate diminishes, so does the stress, until failure ensues. In addition, the thermal fluctuations mediated by random initial velocity generation show a relatively small impact on the number of broken bonds and the stress response during network deformation, though the bond breaking rate can fluctuate more strongly (in Fig. 6c) due to discreteness and randomness in individual bond breakages. The total count of bonds required to rupture the thermoset is approximately 55, around 0.2% of the total bonds or 1.5% of the potentially breakable bond types in the system, which is notably smaller than the number of reactions needed for network gelation.

Strain rate effect

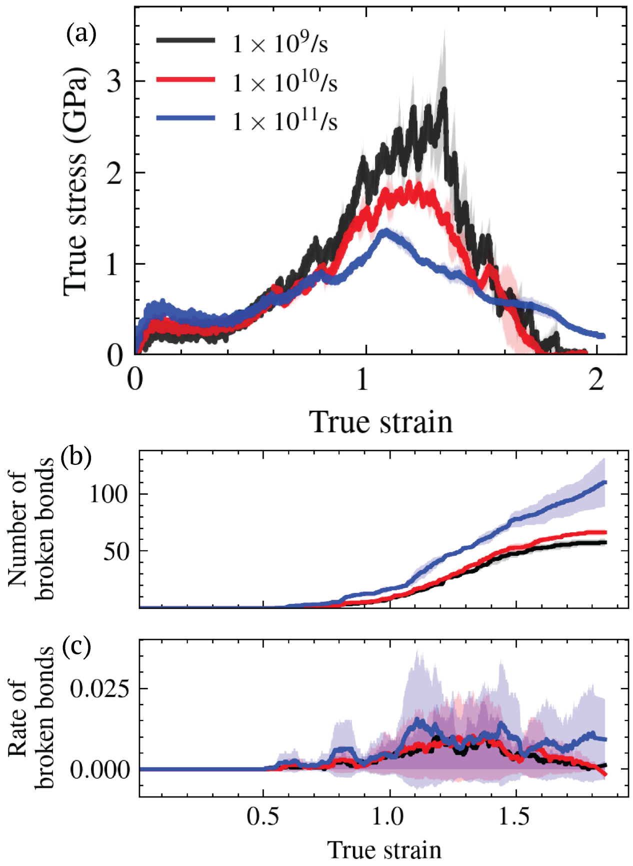

Figure 7: (a) Stress-strain curves of the same initial structure (98% crosslinked) simulated by MLABT with three strain rates at 300 K. (b)-(c) Corresponding number (b) and rate (c) of the bond breaking events during deformation.

Next, we investigate the effect of strain rate on the thermoset fracture behavior. As shown in Fig. 7, distinct from temperature, strain rate not only strongly affects the elastic behavior but also plastic and fracture behaviors. In the elastic regime, a larger strain rate results in higher elastic modulus and yield stress (Fig 7A), in agreement with previous experiments and simulations.[47,48,54–56] Thus, a more noticeable strain softening regime is observed at larger strain rates. More interestingly, the bond breaking occurs more rapidly at higher strain rates, although the characteristic strains for fracture initiation (first bond breaking) and ultimate stress are almost independent of strain rate, as shown in Fig. 7B and C. Consequently, more bonds are broken in deformation with a large strain rate and the resultant ultimate stress is evidently reduced. The reason for this result is that at large strain rates with little stress release, bond breakages do not occur in the most productive way, i.e., breaking apart the network using as few cuts as possible. Hence, the system could remain connected even though a large number of bonds have broken, and as a result, the system exhibits reduced ultimate stress and behaves more ductile at large strain rates, as shown in Fig. 7A. Note that this effect might converge at low strain rates, as the stress is fully relaxed once a bond breaks when the associated timescale is closer to or even longer than the stress field propagation time, which is supported by the smaller difference in the effect when changing the strain rate from 10^{10}/s to 10^{9}/s. Nevertheless, the MLABT results suggest a distinct disparity between effects of strain rate and temperature within the plastic regime. A further demonstration on the breakdown of their superposition is provided in the Supporting Information.

Cooling rate effect

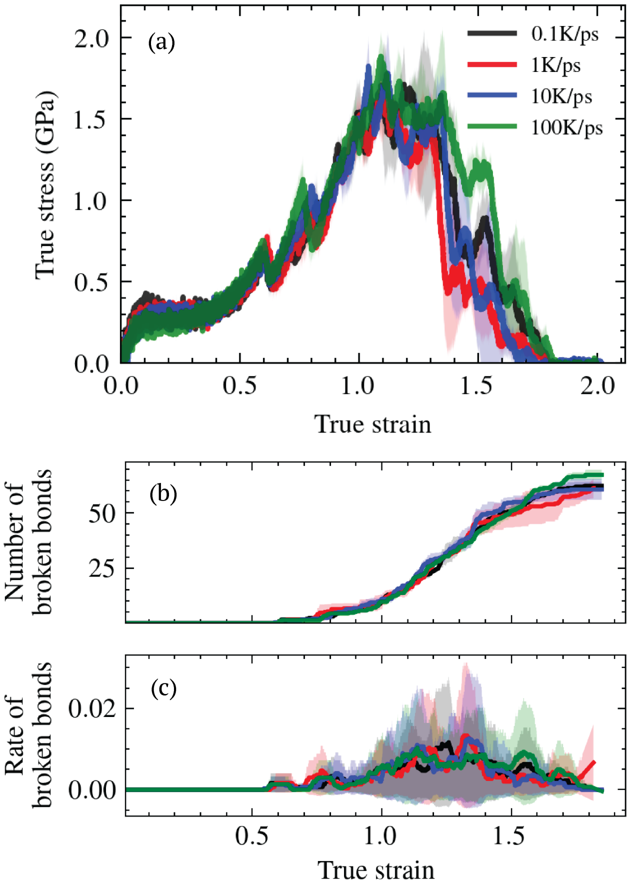

Figure 8: (a) Stress-strain curves of polymer networks with the same topology (98% crosslinked) but generated with different annealing (cooling) rates simulated by MLABT at 300 K and with a strain rate of 10^{10}/s. (b)-(c) Corresponding numbers and rates of the bond breaking events during deformation.

Furthermore, we study the effect of cooling rate in the melt-quenching process for synthesizing glassy thermosets on its bond breaking and fracture behavior by MLABT simulations. As glass is a non-equilibrium state below the glass transition, the cooling rate and the associated timescale for relaxation controls the temperature at which the supercooled liquid loses ergodicity and the amorphous structure becomes "frozen".[57] Specifically, a smaller cooling rate results in a lower glass transition temperature, and thereby, a thermodynamically more stable glass state, i.e., a state located lower in the potential energy landscape. These more stable glasses show distinct atomic structures and materials properties compared to glasses with higher cooling rates, such as higher density and higher mechanical strength.[58,59] Such a cooling rate effect is observed in the elastic regime of thermosets, as shown in Fig. 8A, as the elastic modulus and yield stress increase with reduced cooling rates (details summarized in Table S1 of the Supporting Information). Note that in the simulations, the initial structures have identical bonding topology but only are generated by different cooling rates in melt-quenching simulations initial configurations from 800 K to 300 K prior to deformation. Interestingly, the cooling rate effect does not survive in the plastic behaviors. As strain increases in the strain softening and the initial part of the strain hardening (till strain$\sim$0.6), the difference of stress induced in the elastic regime gradually disappears, suggesting that the effect of the initial glassy structures diminishes. Consequently, the following bond breaking effects and the fracture behavior are independent of the cooling rate, as shown in Fig. 8A-C. These observations can be understood by considering that the cooling rate in general determines the stability of initial glassy structures and therefore controls the elastic behavior (without structural change), however, as strain increases, the strain-induced structural modulation becomes dominant through the covalent bonded network, and the difference in the amorphous strained structures becomes negligible. We should also note that this limited cooling rate effect is an outcome of strong topological constraints of thermosets and the short length of strands used in the simulations.

Effect of crosslinking degree

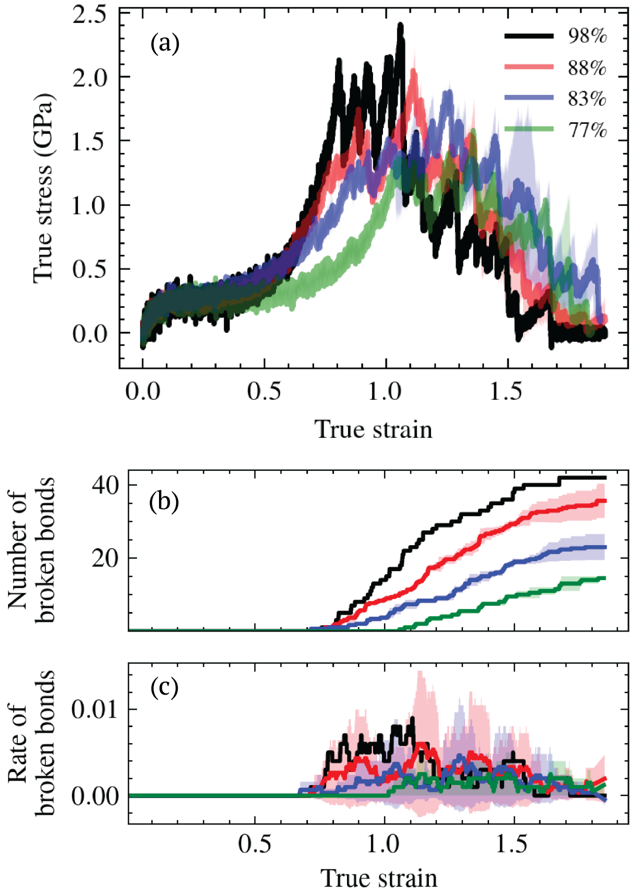

Figure 9: (a) Stress-strain curves simulated by MLABT for structures with different degrees of crosslinking at 300 K and with a strain rate of 10^9/s. (b)-(c) Corresponding numbers and rates of the bond breaking events during deformation.

Finally, we study how the degree of crosslinking affects bond breakage and fracture behavior of thermosets by MLABT simulations. In this work, the crosslinks are formed dynamically in MD simulations of curing reactions, and thus the degree of crosslinking is controlled by how long the curing process runs. As shown in Fig. 9, we prepare four structures with various degrees of crosslinking from 77% to 98% (all gels) and simulate their deformation responses with MLABT at 300 K with a strain rate of 10^9/s. Note that in this work, we do not provide statistically averaged results over various bonding topologies due to computational cost, however, the structures evaluated herein with the four degrees of crosslinking are generated from the same curing reaction in an effort to emphasize the influence of crosslinking degree.

In the system with a higher degree of crosslinking, while the elastic behaviors are subtly changed (due to smaller variations of crosslinking density[60]), the SSC in strain hardening increases more rapidly, and reaches the ultimate stress at a smaller strain value, as shown in Fig. 9A. As the degree of crosslinking decreases, the ultimate stress shifts to a larger strain and the peak value decreases. Specifically, when the crosslinking degree decreases from 98% to 77%, the ultimate stress decreases by roughly 1/3 and the corresponding true strain increases by 1/3. This is generally consistent with results of previous simulations using ReaxFF.[45] This behavior of SSC can be understood from bond breakages in Fig. 9B and C. As expected, fewer bonds are broken in systems with lower crosslinking degrees, consistent with the lower ultimate stress values. Moreover, the dominant bond breaking events, indicated by the maximum rate of bond breaking, occur at larger strains, accounting for the shift of ultimate strain. Depending on specific modification of bonding topology, the strain of fracture initiation might shift such that it is delayed to a larger strain at aa crosslinking degree of 77%. We also note that although the effect of crosslinking degree is revealed here, the relationship between network fracture and network topology is intriguing yet more complex (the variance induced by topology at same degrees of crosslinking is illustrated in Fig. S7 of the Supporting Information), which warrants further investigation.

Discussion

The AL-DFT MLABT framework stands out as a promising approach for accurately modeling thermoset fracture, as bond breaking during fracture is performed at nearly quantum chemical accuracy while computational cost is kept similar to that of classical MD. Specifically, the AL modification of MLABT presents two distinct advantages: i) Integration of AL enables incorporation of diverse training data across thermodynamic and configurational conditions, improving the transferability of models to new conditions. ii) By actively pinpointing the most informative data for training, the AL framework is roughly an order of magnitude more efficient in the generation of training data than our previous MLABT model. This heightened efficiency permits utilization of more accurate (and computationally costly) QM methods, such as DFT.

Importantly, these two advantages of the AL-DFT MLABT framework manifest in the potential generalizability of the methodology across chemical space, an area where other reactive models (e.g. ReaxFF) can struggle. As MLABT leverages widely-accepted classical force-fields, such as OPLS and Amber, as its baseline simulation method, the high performance of common quantities such as density and low strain mechanical response is ensured across a broad chemical space. By adding on a QM-quality bond breaking prediction, mediated by ML, to these classical simulations, the MLABT approach is inherently adaptable to new chemistries for which accurate classical force-fields exist. Moreover, the ML component responsible for bond breaking is also inherently adaptable as it leverages structural representations (SOAP vectors) and is informed by QM calculations, making it agnostic to specific chemistries. This flexibility allows for the efficient adaptation of the MLABT framework to new chemistries, a process further streamlined by the AL approach introduced in this study. In contrast, empirical methods like ReaxFF demand a cumbersome and iterative parameterization process.[61] Transferability of the parameter sets in ReaxFF is a well-known (and expected) deficiency compared to classical FF as the prediction task required is more challenging.[62] In our testing as detailed in the supporting information, MLABT and ReaxFF show generally consistent results in modeling network deformation, particularly on the overall bond breakages. However, without constant computation of charge and bond topology as in ReaxFF, MLABT is computationally more efficient by at least one order of magnitude. Provided these advantages, MLABT may be a suitable and easily implementable alternative to ReaxFF for unparameterized chemistries, especially when a trusted classical FF is already known.

As mentioned in the Methods section, a minor limitation of the current MLABT implementation is the necessity for a small timestep (0.25 fs as opposed to 1 fs commonly used in standard MD simulations). This issue is primarily due to the instability introduced by broken bonds when using classical force fields. However, this drawback can be fixed in future developments of MLABT, either through optimizing the force fields or by appending hydrogens to atoms involved in the bond rupture. Importantly, the MLABT simulations conducted in this study did not employ these modifications, ensuring that our reported outcomes remain free from any influence of mechanoradicals that could otherwise alter network dynamics and subsequent bond breakages. The role of these mechanoradicals in subsequent bond breaking processes is a challenging topic reserved for future works.

The computational efficiency of MLABT relative to other QM-informed bond-breaking approaches also enables a number of crucial insights into the nature of fracture in epoxy thermosets. Specifically, we are able to simulate multiple replicas of our networks both in phase space and topology with DFT accuracy at minimal computational cost. From this added computational sampling, we observe that thermal fluctuations in the fracture process (as mediated by differing initial seeds in the velocity distribution) have a small impact on the resulting fracture process until the point of material failure. Although the exact bond breaking sites can differ due to randomness in the initial velocity distribution, the variations in the SSC and bond breaking are small during strain hardening. After the ultimate stress, these variations become larger due to accumulated topological differences.

Importantly, the MLABT results demonstrate that the polymer network topology plays a more important role in controlling fracture behavior than fine details of the polymer's glassy structure. Specifically, although the elastic mechanical properties depend on the cooling rate in the melt-quenching of thermosets, the plastic properties and bond breakages are independent of it (Fig. 8), suggesting that the influence of glassy structure blurs due to strain-induced rearrangement beyond the elastic regime. On the other hand, the degree of crosslinking (Fig. 9) and the topology variation at the same degree (Fig. S7) primarily determine the fracture behavior, such as the ultimate strain/stress and the fraction of broken bonds. This again confirms the importance of topological perspective for understanding the fracture mechanisms of polymer networks.[7,9,63]

Lastly, our results show that time and temperature, the superposition of which is often posited to understand the viscoelastic behaviors of polymers, play different roles in the network fracture. Specifically, temperature does not affect bond breakages but slightly changes the resultant stress due to the local relaxation of glassy structures. However, the strain rate can strongly determine the strain-stress behavior by influencing the rate of bond breakages. At large strain rates, the strain-induced structural evolution can propagate faster than the release of stress via network topology immediately after bond scission. As a consequence, additional "unnecessary" bonds break, resulting in a more fragile network with lower ultimate stress. Note that although this physical scenario seems reasonable from the atomistic perspective (also supported by the ReaxFF results in Fig. S6 of the Supporting Information), it cannot explain the general trend observed in experiments that a decrease in temperature or an increase in strain rate leads to more brittle fracture behaviors. The reason is that in experiments where deformation is much slower compared with stress release, the bond breakages can occur in the most efficient way through networks and the difference in fraction of broken bonds is tiny when varying deformation rates. This is consistent with the simulation results in Fig. 7b where the differences become smaller when further decreasing the strain rate, suggesting a possible convergence at strain rates close to those used in experiments. As such, the ultimate stress on experimental timescales will be dominated by stress relaxation due to structural reorganization. This is confirmed by our simulation results that show the stress after structural optimization at each deformation step is much lower than that without optimization, as shown in Fig. S9 in the Supporting Information. In this context, the strain rate effect reported in the MLABT simulations is mainly due to bond breaking events rather than long-term structural relaxation.

More importantly, the discrepancy with experiments is reflected in the long-recognized anomaly of ductile fracture in the molecular modeling of thermosets,[64,65] a phenomenon whose root cause is still under debate.[66] Given that the MLABT method offers QM-level accuracy in bond breaking, yet still demonstrates ductile fracture, and considering the revealed atomistic strain-rate effect is unlikely to contribute to a brittle fracture (when extrapolated at experimental strain rates), modeling at larger length scales is imperative to bridge the gap between experiments and simulations. One reason is that for typical simulation length scales, the crosslinking density ($\sim$0.1 mole/cm^3) significantly exceeds the experimental values (0.001-0.01 mole/cm^3).[67,68] This discrepancy hinders the formation of polymer chain entanglements in current molecular dynamics simulations of thermosets. Another hypothesis is related to larger-scale structures, such as air pockets or pores.

Air pockets or pores could potentially expedite the fracture towards the elastic regime and exhibit a distinct temperature dependence through micro-structural reorganization.[69–71] One piece of supporting evidence is that in experiments of thermoset films where length-scales are limited below micrometers, the elongation can exceed 100% before failure,[72] in better alignment with the behaviors in the simulations. These factors also contribute to the ultimate strength at fracture of the material in the present simulations being greater than those in experiments. More detailed discussions are provided in the Supporting Information. Nevertheless, we believe that incorporating larger length-scale structures in modeling is a promising direction for resolving these discrepancies, and MLABT, owing to its efficient and accurate treatment of bond breaking, is well-suited for extending to larger length-scales in diverse environments, including composites and interfaces.

Conclusion

We have combined AL with MLABT to create a framework for the atomistic modeling of thermoset fracture with DFT-accuracy and classical FF cost. This integration offers heightened predictive accuracy across thermodynamic and configurational space while simultaneously improving efficiency during model training. Moreover, the adaptability of MLABT with AL enables researchers to explore polymer fracture across diverse chemistries, provided a suitable classical FF, avoiding the pitfalls of cumbersome parameterization in other reactive methods. We have applied the MLABT framework to understand the molecular determinants of fracture in polymer networks, finding that the network topology largely dictates fracture behaviors, while the intrinsic stability of the glassy structures has little influence. Interestingly, strain rate, rather than temperature, predominantly impacts network fracture at atomic scale, deviating from conventional paradigms in elasticity. Looking ahead, MLABT-based models of thermoset fracture form a potentially useful basis for the establishment of chemically generalizable bond-breaking models applicable to a diverse set of polymer chemistries while leveraging the vast array of existing, and high-accuracy, classical force-fields.

Supporting Information

Uncertainty and convergence of the AL MLABT model, effect of MLABT bond scanning frequency on failure, comparison of MLABT with a simple model based on bond lenths, comparison of MLABT with ReaxFF, effect of network topology on fracture, time-temperature superposition in polymer network fracture, discussion on discrepancies with experiments.

Acknowledgement

This material is based upon work supported by the National Science Foundation Chemical Theory, Models, and Computation division under award CHE-2154916. This work used Bridges-2 at the Pittsburgh Supercomputing Center through allocation CHE230055 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

References

(1) Jayan, J. S.; Appukuttan, S.; Wilson, R.; Joseph, K.; George, G.; Oksman, K. In *Fiber Reinforced Composites*; Joseph, K., Oksman, K., George, G., Wilson, R., Appukuttan, S., Eds.; Woodhead Publishing Series in Composites Science and Engineering; Woodhead Publishing, 2021; pp 1–24.

(2) Auvergne, R.; Caillol, S.; David, G.; Boutevin, B.; Pascault, J.-P. Biobased Thermosetting Epoxy: Present and Future. *Chem. Rev.* **2014**, *114*, 1082–1115.

(3) Jin, F.-L.; Li, X.; Park, S.-J. Synthesis and Application of Epoxy Resins: A Review. *Journal of Industrial and Engineering Chemistry* **2015**, *29*, 1–11.

(4) Post, W.; Susa, A.; Blaauw, R.; Molenveld, K.; Knoop, R. J. I. A Review on the Potential and

Limitations of Recyclable Thermosets for Structural Applications. *Polymer Reviews* **2020**, *60*, 359–388.

(5) Gioia, C.; Lo Re, G.; Lawoko, M.; Berglund, L. Tunable Thermosetting Epoxies Based on Fractionated and Well-Characterized Lignins. *J. Am. Chem. Soc.* **2018**, *140*, 4054–4061.

(6) Rottler, J. Fracture in Glassy Polymers: A Molecular Modeling Perspective. *J. Phys.: Condens. Matter* **2009**, *21*, 463101.

(7) Barney, C. W.; Ye, Z.; Sacligil, I.; McLeod, K. R.; Zhang, H.; Tew, G. N.; Riggleman, R. A.; Crosby, A. J. Fracture of Model End-Linked Networks. *Proceedings of the National Academy of Sciences* **2022**, *119*, e2112389119.

(8) Wang, S.; Panyukov, S.; Craig, S. L.; Rubinstein, M. Contribution of Unbroken Strands to the Fracture of Polymer Networks. *Macromolecules* **2023**, *56*, 2309–2318.

(9) Stevens, M. J. Interfacial Fracture between Highly Cross-Linked Polymer Networks and a Solid Surface: Effect of Interfacial Bond Density. *Macromolecules* **2001**, *34*, 2710–2718.

(10) Sbrescia, S.; Ju, J.; Creton, C.; Engels, T.; Seitz, M. Effect of Temperature, Rate, and Molecular Weight on the Failure Behavior of Soft Block Copoly(Ether–Ester) Thermoplastic Elastomers. *Soft Matter* **2023**, *19*, 5127–5141.

(11) Ortiz, C.; Kim, R.; Rodighiero, E.; Ober, C. K.; Kramer, E. J. Deformation of a Polydomain, Liquid Crystalline Epoxy-Based Thermoset. *Macromolecules* **1998**, *31*, 4074–4088.

(12) Glotzer, S. C.; Paul, W. Molecular and Mesoscale Simulation Methods for Polymer Materials. *Annual Review of Materials Research* **2002**, *32*, 401–436.

(13) Wu, C.; Xu, W. Atomistic Molecular Modelling of Crosslinked Epoxy Resin. *Polymer* **2006**, *47*, 6004–6009.

(14) Zhong, M.; Wang, R.; Kawamoto, K.; Olsen, B. D.; Johnson, J. A. Quantifying the Impact of Molecular Defects on Polymer Network Elasticity. *Science* **2016**, *353*, 1264–1268.

(15) Odegard, G. M.; Jensen, B. D.; Gowtham, S.; Wu, J.; He, J.; Zhang, Z. Predicting Mechanical Response of Crosslinked Epoxy Using ReaxFF. *Chemical Physics Letters* **2014**, *591*, 175–178.

(16) Li, C.; Strachan, A. Molecular Scale Simulations on Thermoset Polymers: A Review. *Journal of Polymer Science Part B: Polymer Physics* **2015**, *53*, 103–122.

(17) Hsu, Y.-C.; Yu, C.-H.; Buehler, M. J. Using Deep Learning to Predict Fracture Patterns in Crystalline Solids. *Matter* **2020**, *3*, 197–211.

(18) Buehler, M. J. Modeling Atomistic Dynamic Fracture Mechanisms Using a Progressive Transformer Diffusion Model. *Journal of Applied Mechanics* **2022**, *89*.

(19) Wang, S.; Hu, Y.; Kouznetsova, T. B.; Sapir, L.; Chen, D.; Herzog-Arbeitman, A.; Johnson, J. A.; Rubinstein, M.; Craig, S. L. Facile Mechanochemical Cycloreversion of Polymer Cross-Linkers Enhances Tear Resistance. *Science* **2023**, *380*, 1248–1252.

(20) Zhao, X.; Chen, X.; Yuk, H.; Lin, S.; Liu, X.; Parada, G. Soft Materials by Design: Unconventional Polymer Networks Give Extreme Properties. *Chem. Rev.* **2021**, *121*, 4309–4372.

(21) Barr, S. A.; Kedziora, G. S.; Ecker, A. M.; Moller, J. C.; Berry, R. J.; Breitzman, T. D. Bond Breaking in Epoxy Systems: A Combined QM/MM Approach. *J. Chem. Phys.* **2016**, *144*, 244904.

(22) Yu, Z.; Jackson, N. E. Machine Learning Quantum-Chemical Bond Scission in Thermosets under Extreme Deformation. *Applied Physics Letters* **2023**, *122*, 211906.

(23) Lookman, T.; Balachandran, P. V.; Xue, D.; Yuan, R. Active Learning in Materials Science with Emphasis on Adaptive Sampling Using Uncertainties for Targeted Design. *npj Comput Mater* **2019**, *5*, 1–17.

(24) Prince, M. Does Active Learning Work? A Review of the Research. *Journal of Engineering Education* **2004**, *93*, 223–231.

(25) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less Is More: Sampling Chemical Space with Active Learning. *The Journal of Chemical Physics* **2018**, *148*, 241733.

(26) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent Advances and Applications of Machine Learning in Solid-State Materials Science. *npj Comput Mater* **2019**, *5*, 1–36.

(27) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(28) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in ’t Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comp. Phys. Comm.* **2022**, *271*, 108171.

(29) Grandbois, M.; Beyer, M.; Rief, M.; Clausen-Schaumann, H.; Gaub, H. E. How Strong Is a Covalent Bond? *Science* **1999**, *283*, 1727–1730.

(30) Garcia, F. G.; Soares, B. G.; Pita, V. J. R. R.; Sánchez, R.; Rieumont, J. Mechanical Properties of Epoxy Networks Based on DGEBA and Aliphatic Amines. *Journal of Applied Polymer Science* **2007**, *106*, 2047–2055.

(31) Vashisth, A.; Ashraf, C.; Bakis, C. E.; van Duin, A. C. T. Effect of Chemical Structure on Thermo-Mechanical Properties of Epoxy Polymers: Comparison of Accelerated ReaxFF Simulations and Experiments. *Polymer* **2018**, *158*, 354–363.

(32) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* **2013**, *87*, 184115.

(33) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing Molecules and Solids across Structural and Alchemical Space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.

(34) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Federici Canova, F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. DScribe: Library of Descriptors for Machine Learning in Materials Science. *Computer Physics Communications* **2020**, *247*, 106949.

(35) Chang, C.-C.; Lin, C.-J. LIBSVM: A Library for Support Vector Machines. *ACM transactions on intelligent systems and technology (TIST)* **2011**, *2*, 1–27.

(36) Boerner, T. J.; Deems, S.; Furlani, T. R.; Knuth, S. L.; Towns, J. *Practice and Experience in Advanced Research Computing*; 2023; pp 173–176.

(37) Cohn, D. A.; Ghahramani, Z.; Jordan, M. I. Active Learning with Statistical Models. *Journal of artificial intelligence research* **1996**, *4*, 129–145.

(38) Settles, B. Active Learning Literature Survey. **2009**,

(39) Settles, B. From Theories to Queries: Active Learning in Practice. Active Learning and Experimental Design Workshop In Conjunction with AISTATS 2010. 2011; pp 1–18.

(40) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.

(41) Schohn, G.; Cohn, D. Less Is More: Active Learning with Support Vector Machines. ICML. 2000; p 6.

(42) Kremer, J.; Steenstrup Pedersen, K.; Igel, C. Active Learning with Support Vector Machines. *WIREs Data Mining and Knowledge Discovery* **2014**, *4*, 313–326.

(43) Platt, J., et al. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Advances in large margin classifiers* **1999**, *10*, 61–74.

(44) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA Quantum Chemistry Program Package. *J. Chem. Phys.* **2020**, *152*, 224108.

(45) Meng, Z.; Bessa, M. A.; Xia, W.; Kam Liu, W.; Keten, S. Predicting the Macroscopic Fracture Energy of Epoxy Resins from Atomistic Molecular Simulations. *Macromolecules* **2016**, *49*, 9474–9483.

(46) Kubát, M.; Matwin, S. Addressing the Curse of Imbalanced Training Sets: One-sided Selection. International Conference on Machine Learning. 1997.

(47) Konrad, J.; Meißner, R. H.; Bitzek, E.; Zahn, D. A Molecular Simulation Approach to Bond Reorganization in Epoxy Resins: From Curing to Deformation and Fracture. *ACS Polym. Au* **2021**, *1*, 165–174.

(48) Moller, J. C.; Barr, S. A.; Schultz, E. J.; Breitzman, T. D.; Berry, R. J. Simulation of Fracture Nucleation in Cross-Linked Polymer Networks. *JOM* **2013**, *65*, 147–167.

(49) Li, C.; Medvedev, G. A.; Lee, E.-W.; Kim, J.; Caruthers, J. M.; Strachan, A. Molecular Dynamics Simulations and Experimental Studies of the Thermomechanical Response of an Epoxy Thermoset Polymer. *Polymer* **2012**, *53*, 4222–4230.

(50) Li, C.; Strachan, A. Molecular Dynamics Predictions of Thermal and Mechanical Properties of Thermoset Polymer EPON862/DETDA. *Polymer* **2011**, *52*, 2920–2928.

(51) Yoshimoto, K.; Jain, T. S.; Workum, K. V.; Nealey, P. F.; de Pablo, J. J. Mechanical Heterogeneities in Model Polymer Glasses at Small Length Scales. *Phys. Rev. Lett.* **2004**, *93*, 175501.

(52) Riggleman, R. A.; Lee, H.-N.; Ediger, M. D.; de Pablo, J. J. Free Volume and Finite-Size Effects in a Polymer Glass under Stress. *Phys. Rev. Lett.* **2007**, *99*, 215501.

(53) Pink, E.; Campbell, J. D. The Low-Temperature Macro Deformation of an Epoxide Resin. *Materials Science and Engineering* **1974**, *15*, 187–194.

(54) Mayr, A. E.; Cook, W. D.; Edward, G. H. Yielding Behaviour in Model Epoxy Thermosets — I. Effect of Strain Rate and Composition. *Polymer* **1998**, *39*, 3719–3724.

(55) Fan, J.; Anastassiou, A.; Macosko, C. W.; Tadmor, E. B. Molecular Dynamics Predictions of Thermomechanical Properties of an Epoxy Thermosetting Polymer. *Polymer* **2020**, *196*, 122477.

(56) Tamrakar, S.; Ganesh, R.; Sockalingam, S.; Haque, B. Z.; Gillespie, J. W. Experimental Investigation of Strain Rate and Temperature Dependent Response of an Epoxy Resin Undergoing Large Deformation. *J. dynamic behavior mater.* **2018**, *4*, 114–128.

(57) Ediger, M. D.; Angell, C. A.; Nagel, S. R. Supercooled Liquids and Glasses. *J. Phys. Chem.* **1996**, *100*, 13200–13212.

(58) Barrat, J.-L.; Baschnagel, J.; Lyulin, A. Molecular Dynamics Simulations of Glassy Polymers. *Soft Matter* **2010**, *6*, 3430–3446.

(59) Yu, Z.; Liu, Q.; Szlufarska, I.; Wang, B. Structural Signatures for Thermodynamic Stability in Vitreous Silica: Insight from Machine Learning and Molecular Dynamics Simulations. *Phys. Rev. Materials* **2021**, *5*, 015602.

(60) Shokuhfar, A.; Arab, B. The Effect of Cross Linking Density on the Mechanical Properties and Structure of the Epoxy Polymers: Molecular Dynamics Simulation. *J Mol Model* **2013**, *19*, 3719–3731.

(61) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.

(62) Liu, L.; Liu, Y.; Zybin, S. V.; Sun, H.; Goddard, W. A. I. ReaxFF-lg: Correction of the ReaxFF Reactive Force Field for London Dispersion, with Applications to the Equations of State for Energetic Materials. *J. Phys. Chem. A* **2011**, *115*, 11016–11022.

(63) Lin, S.; Zhao, X. Fracture of Polymer Networks with Diverse Topological Defects. *Phys. Rev. E* **2020**, *102*, 052503.

(64) Wu, X.; El-Awady, J. A. In *Integrated Computational Materials Engineering (ICME): Advancing Computational and Experimental Methods*; Ghosh, S., Woodward, C., Przybyla, C., Eds.; Springer International Publishing: Cham, 2020; pp 267–296.

(65) Littell, J. D.; Ruggeri, C. R.; Goldberg, R. K.; Roberts, G. D.; Arnold, W. A.; Binienda, W. K. Measurement of Epoxy Resin Tension, Compression, and Shear Stress–Strain Curves over a Wide Range of Strain Rates Using Small Test Specimens. *Journal of Aerospace Engineering* **2008**, *21*, 162–173.

(66) Bukowski, C.; Zhang, T.; Riggleman, R. A.; Crosby, A. J. Load-Bearing Entanglements in Polymer Glasses. *Science Advances* **2021**, *7*, eabg9763.

(67) Chen, J.-S.; Ober, C. K.; Poliks, M. D.; Zhang, Y.; Wiesner, U.; Cohen, C. Controlled Degradation of Epoxy Networks: Analysis of Crosslink Density and Glass Transition Temperature Changes in Thermally Reworkable Thermosets. *Polymer* **2004**, *45*, 1939–1950.

(68) Zhao, S.; Abu-Omar, M. M. Renewable Epoxy Networks Derived from Lignin-Based Monomers: Effect of Cross-Linking Density. *ACS Sustainable Chem. Eng.* **2016**, *4*, 6082–6089.

(69) Schichtel, J. J.; Chattopadhyay, A. Modeling Thermoset Polymers Using an Improved Molecular Dynamics Crosslinking Methodology. *Computational Materials Science* **2020**, *174*, 109469.

(70) Bay, R. K.; Zhang, T.; Shimomura, S.; Ilton, M.; Tanaka, K.; Riggleman, R. A.; Crosby, A. J. Decoupling the Impact of Entanglements and Mobility on the Failure Properties of Ultrathin Polymer Films. *Macromolecules* **2022**, *55*, 8505–8513.

(71) Wu, X.; Aramoon, A.; El-Awady, J. A. Hierarchical Multiscale Approach for Modeling the Deformation and Failure of Epoxy-Based Polymer Matrix Composites. *J. Phys. Chem. B* **2020**, *124*, 11928–11938.

(72) Boarino, A.; Charmillot, J.; Figueirêdo, M. B.; Le, T. T. H.; Carrara, N.; Klok, H.-A. Ductile, High-Lignin-Content Thermoset Films and Coatings. *ACS Sustainable Chem. Eng.* **2023**, *11*, 16442–16452.