

On-the-Fly Machine-Learned Force Fields for High-Fidelity Polymer Glass Transition Simulations

Ashutosh Srivastava, Sakshi Agarwal, Shivank Shukla, Harikrishna Sahu, and Rampi Ramprasad*



Cite This: <https://doi.org/10.1021/acs.jcim.6c02116>



Read Online

ACCESS |



Metrics & More

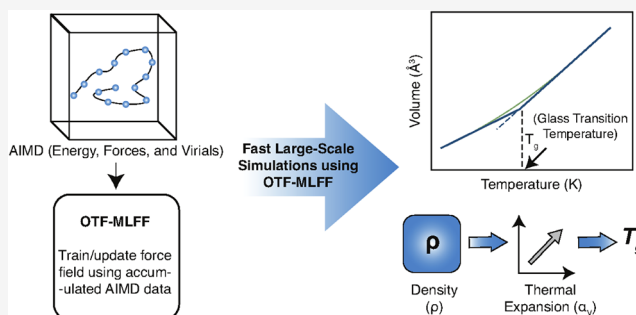


Article Recommendations



Supporting Information

ABSTRACT: Predicting polymer glass transition temperatures (T_g) with first-principles fidelity has long remained out of reach, as cooling multithousand-atom systems over a broad temperature range at acceptable rates exceeds the computational limits of ab initio molecular dynamics (AIMD). Here we employ a hybrid scheme that merges AIMD with accelerated on-the-fly (OTF) machine-learned force-field (MLFF) construction, enabling T_g prediction at near quantum-mechanical accuracy with near-classical computational cost. The OTF protocol to construct MLFFs adaptively triggers first-principles calculations only when newly encountered configurations lie outside the current model's domain of confidence, allowing robust, parameter-free MLFFs to be built from merely ~ 1000 AIMD-sampled configurations per polymer. These MLFFs are then utilized to perform long-time cooling simulations on amorphous supercells containing several thousand atoms. Applied across 12 polymers spanning aromatic, aliphatic, heteroatomic, and branched chemistries, the method yields predictions in excellent accord with experiment while reducing computational cost by approximately 6 orders of magnitude relative to AIMD. This work establishes a new paradigm for predictive polymer modeling, demonstrating that OTF-MLFFs provide a generalizable, accurate, and scalable route to simulating the thermophysical behavior of complex disordered materials at near quantum-mechanical fidelity.



INTRODUCTION

High-fidelity simulations of complex macromolecular systems remain one of the central challenges in computational materials science. Many phenomena of practical and fundamental interest, including structural reorganizations in dense or disordered environments, thermophysical transitions, and collective fluctuations, unfold over nanoseconds or longer and involve thousands of atoms.^{1–5} Polymers are emblematic of this challenge: their behavior is governed by slow configurational rearrangements, broad distributions of local environments, and pronounced dynamic heterogeneity that cannot be captured within the spatial or temporal windows accessible to conventional ab initio molecular dynamics (AIMD). Although AIMD provides unparalleled quantum-mechanical fidelity, its steep computational scaling confines simulations to small unit cells and/or short trajectories, rendering it inadequate for capturing the full physics of polymeric materials.

In recent years, machine-learned force fields (MLFFs) have emerged as a transformative pathway for accelerating atomistic simulations toward quantum accuracy. Frameworks based on pretrained or transferable architectures, including NequIP,⁶ Allegro,⁷ MACE,⁸ and DeepMD,⁹ as well as kernel-based approaches, have significantly reduced the amount of system-specific training data required through transfer learning and fine-tuning strategies. Nevertheless, accurately describing

polymeric systems remains challenging because their structural diversity, long-time scale dynamics, and complex phase behavior are often underrepresented in the data sets used to develop foundation models. As a result, additional polymer-specific training data and fine-tuning are frequently required to achieve reliable predictions. For example, recent benchmarking of MACE-OFF demonstrated limited transferability for several polymers, including Polyacrylonitrile (PAN), Polyethylene terephthalate (PET), and Poly[thio bis(4-phenyl)carbonate] (PTPC), while simulations of Polychlorotrifluoroethylene (PCTFE) were found to be unstable.^{8,10} Consequently, although foundation models substantially reduce the cost of force-field development, obtaining highly accurate MLFFs for polymer simulations can still require considerable data curation, model refinement, and computational resources. These issues highlight the challenges in achieving both stability and physical fidelity in MLFF-driven polymer simulations. For polymers, where a complex hierarchy of primary and secondary

Received: June 29, 2026

Revised: August 19, 2026

Accepted: August 21, 2026

interactions coexists and where configurational diversity is vast, such data demands and training burdens have impeded widespread application.

Hybrid QM/ML strategies offer an appealing alternative by invoking first-principles calculations only when needed. Among these, on-the-fly (OTF) MLFF learning stands out for its adaptivity, sample efficiency, and elimination of pregenerated training sets.¹¹ In the OTF-MLFF paradigm, a simulation begins with a short AIMD segment that seeds an initial MLFF. As the trajectory proceeds, a Bayesian uncertainty estimator identifies configurations insufficiently represented in the model. Only in these cases is a new first-principles calculation triggered,¹² after which the training set is expanded and the MLFF refined in real time (Figure 1a). As configurational

simulating the glass transition phenomenon in polymers using standalone MLFFs such as Vivace¹⁰ and MACE-OFF,⁸ while promising, also highlight key challenges, likely rooted in the careful selection of the training data. These approaches either fail to reveal the glass transition with sufficient clarity or fail entirely for certain polymer chemistries (see Figures S7 and S8 in the Supporting Information of the work¹⁰).

Here, we present the first generalizable and scalable OTF-MLFF^{12,19} framework capable of simulating polymer T_g with ab initio fidelity across a broad chemical space, enabled by the capability implemented in the Vienna Ab-initio Simulation Package (VASP).^{20,21} We evaluate 12 polymers spanning aromatic, aliphatic, heteroatomic, and branched architectures. For each system, a MLFF is generated from a 200 ps NPT equilibration at elevated temperatures (600–800 K), using only ~ 1000 quantum-sampled configurations. These MLFFs are then deployed to equilibrate large amorphous supercells containing several thousand atoms and to perform gradual cooling from the melt to cryogenic temperatures (Figure 1b). From the resulting volume–temperature profiles, T_g , density, and thermal expansion coefficients are extracted. Across all polymers, the predicted quantities exhibit excellent agreement with experiment while reducing the computational cost by roughly 6 orders of magnitude relative to a full AIMD treatment.

More broadly, this work establishes a new paradigm for physics-driven predictive polymer modeling, delivering near-quantum-mechanical accuracy at near-classical computational expense. By making high-fidelity, large-scale simulations of polymers tractable on modest computational resources, this approach unlocks unprecedented opportunities to interrogate complex polymer phenomena.

COMPUTATIONAL METHODOLOGY

Structure Generation

All the initial polymer structures were generated using an updated Polymer Structure Predictor (PSP).²² As the RDKit²³ often fails to generate three-dimensional (3D) structures for large or highly branched polymers, its distance–geometry–based embedding algorithm struggles to satisfy competing spatial constraints, resulting in convergence issues and steric clashes during structure generation. To address these challenges, PSP was enhanced to first construct and optimize 3D geometries of monomer and end-cap units individually, followed by their iterative assembly into polymer chains with periodic relaxation using the universal force field (UFF), thereby yielding robust and physically realistic initial geometries without embedding failures. For the construction of amorphous polymer models, PSP employs PACKMOL²⁴ to pack polymer chains within a simulation box while allowing rigid-body rotations and translations; however, this procedure often results in systems with densities lower than the target value. To overcome this limitation, PSP now incorporates an equilibration stage in LAMMPS²⁵ using the GAFF2 force field and the 21-step relaxation protocol proposed by Abbott et al.,²⁶ enabling efficient structural relaxation and equilibration of the amorphous models at the desired temperature.

Molecular Dynamics Details

The generated structures were then equilibrated, and a force-field was simultaneously trained using the OTF approach^{12,27,28} as implemented in VASP (version 6.5.1).^{20,21} The OTF learning approach is a hybrid approach of first-

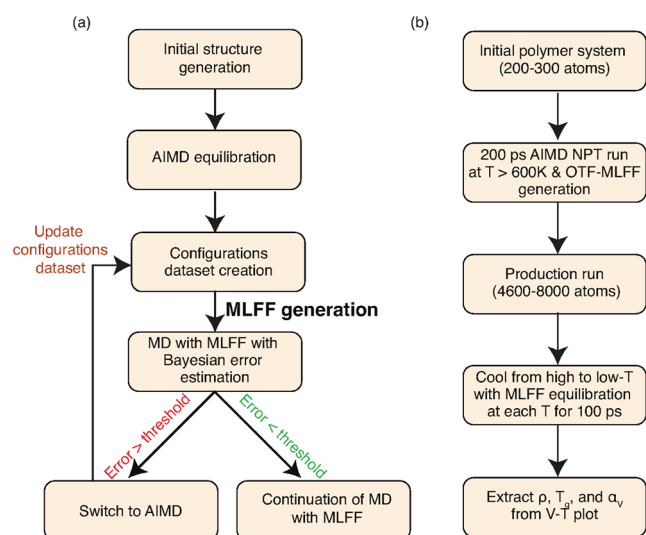


Figure 1. Adaptive on-the-fly machine-learning force-field (OTF-MLFF) framework for polymer thermophysical prediction. (a) Stepwise workflow for generating MLFFs, beginning with initial AIMD equilibration, Bayesian error estimation, automated data set expansion, and real-time force-field refinement. (b) Workflow for predicting polymer glass transition temperature (T_g) and subsequent long-time ML-driven cooling trajectories enabling high-resolution volume–temperature (V – T) analysis to obtain density (ρ), T_g , and volume thermal expansion coefficient (α_v).

diversity increases during the run, the frequency of AIMD evaluations rapidly diminishes, and the force field becomes progressively more robust. The result is a parameter-free, ready-for-use, and physically grounded MLFF that approaches AIMD accuracy (within the domain of the original simulations) while operating near-classical computational cost.

Despite these advances, predicting the glass transition temperature (T_g) of polymers with first-principles fidelity has remained out of reach. T_g marks a fundamental thermophysical transformation, a second-order transition separating rubbery and glassy regimes, and dictates mechanical performance, thermal stability, and processability.^{13–16} Accurate T_g estimation requires slow cooling from the melt, precise detection of subtle volumetric discontinuities, and ample sampling of heterogeneous dynamics, conditions that challenge the computational capabilities of AIMD.^{14,16–18} Classical molecular dynamics can perform long cooling trajectories but relies on empirical force fields whose transferability across polymer families, chemistries, and architectures may be unreliable or is sometimes infeasible. Recent strides in

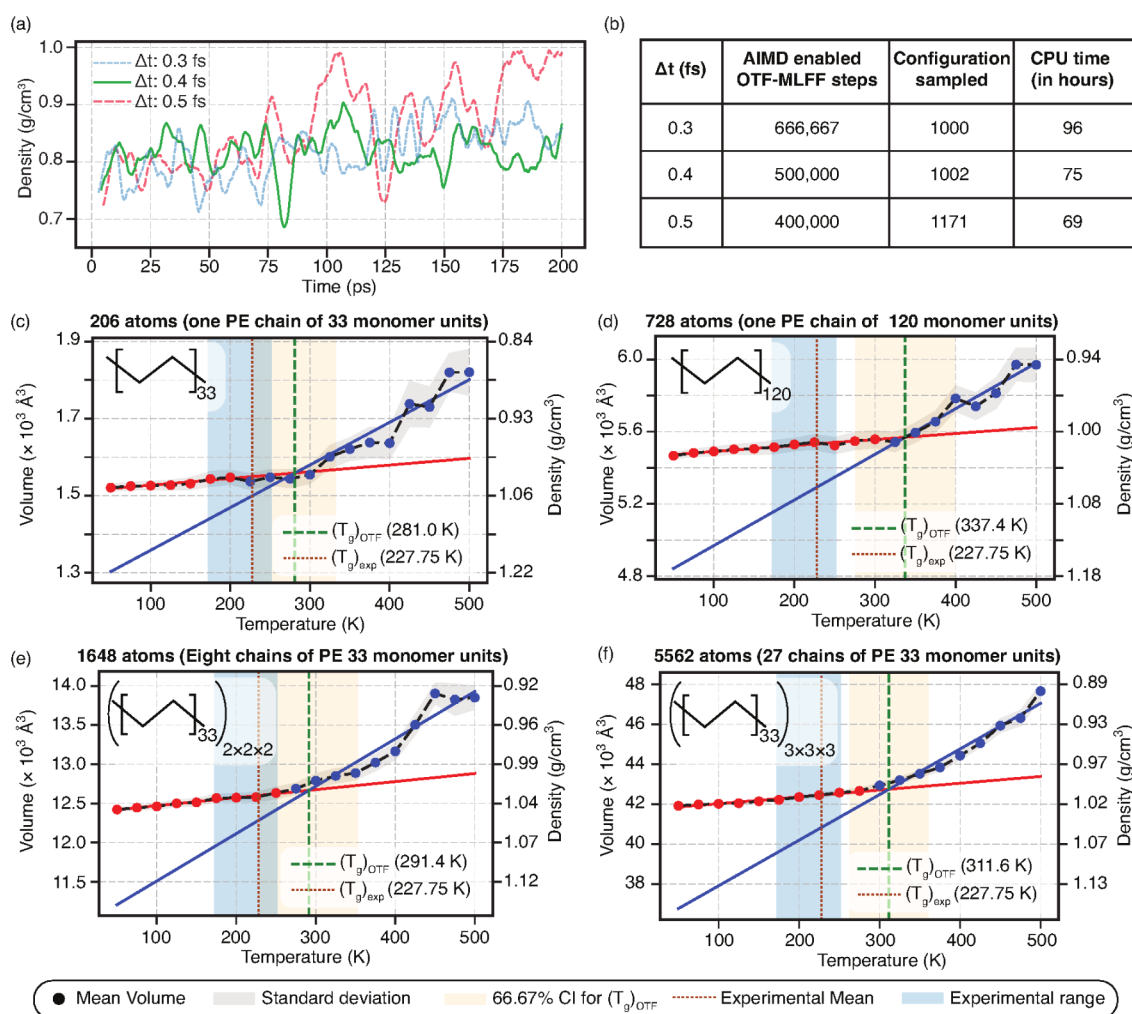


Figure 2. Thermodynamic behavior and computational performance of machine-learned force-field simulations across system sizes. (a) Density as a function of simulation time for different integration step sizes ($\Delta t = 0.3, 0.4$, and 0.5 fs) at 600 K. (b) Summary of time step selection, number of AIMD-enabled OTF-MLFF steps required for a 200 ps trajectory, configurations sampled, and total CPU time (hours). (c–f) Volume (density)–temperature relations for polyethylene systems containing 206, 728, 1648, and 5562 atoms constructed from multiple chains of varying lengths.

principles and machine learning. First-principles calculations were only executed when a new configuration was encountered for which the Bayesian error exceeds the threshold and which was not sampled earlier during the ab initio molecular dynamics (AIMD) run.¹² Similarly, for the configurations that were very close to the configurations already sampled, the approach uses a machine learning algorithm to predict the first-principles output parameters such as forces, energies, and stresses. In this way, we collect all the distinct configurations along with the atomic positions and forces, which are then utilized to build the machine learning force field. OTF-AIMD simulations have been performed using an isothermal–isobaric (NPT) ensemble, allowing the cell shape and volume to relax during dynamics. The Langevin thermostat^{29,30} is employed to maintain temperature and pressure, utilizing the Parinello–Rahman algorithm.^{31,32} The training simulations have been performed for 200 ps (500,000 steps) with an optimized step size of 0.4 fs for polymeric structures containing lighter elements. To facilitate reasonable phase-space sampling, the initial temperature was maintained at 600 K (800 K for polymers with $T_g \geq 400$ K), and the final temperature was set 30% higher, 780 K (1040 K). The pressure was kept at 1 atm during the run.

Electronic Structure Calculation Details

The plane-wave cutoff energy for the electronic charge density expression was set to 500 eV, and the overall precision of the numerical integration was set to *Accurate* to obtain reliable forces and stresses during the run. Interchain van-der Waals interactions were captured using the DFT-D3 scheme³³ with Becke–Johnson damping.³⁴ The partial electronic occupancies were treated using Fermi-smearing with a smearing width of 0.1 eV. The exchange–correlation effects were treated by projector-augmented wave (PAW) pseudopotentials^{35,36} using Predew–Burke–Ernzerhof (PBE) exchange–correlation functional under the generalized gradient approximation (GGA). A $1 \times 1 \times 1$ Γ -centered k -mesh was used to sample the Brillouin zone. The electronic minimization was performed using a mixture of the blocked-Davidson and Residual Minimization Method with Direct Inversion in the Iterative Subspace (RMM-DIIS) algorithms,^{21,37,38} with a convergence criterion of 0.00001 eV during the self-consistent loop.

High-to-Low Temperature Cooling Details

The constructed MLFFs for each polymer were subsequently applied to the corresponding larger polymer structures, each having ~ 20 – 32 polymer chains (see SI-Table S2 for details).

Each system was further equilibrated using the respective MLFFs, starting from the training temperature and cooling down to 50 K in 25 K intervals. At each temperature step, simulations were run for 100 ps, and the final 50 ps of each trajectory was used to construct the volume–temperature (VT) curve. To optimize storage efficiency without compromising accuracy, volume data were recorded every ten steps and used to calculate the average and standard deviation at each temperature. The resulting VT curves were fitted with two linear segments, and the intersection point of these lines was taken as the T_g of the corresponding polymer. Additionally, an error-propagation analysis was performed to estimate the 66.67% confidence interval (CI) of T_g , based on the uncertainties in the fitted slopes (refer to the [Supporting Information](#) for details).

RESULTS AND DISCUSSION

Establishing the Approach Using Polyethylene

In AIMD simulations, the integration time step (Δt) is a critical parameter that controls both accuracy and numerical stability. Larger time steps reduce computational cost but risk significant errors in energies and forces, while smaller time steps improve fidelity at the expense of substantially higher computational effort. This balance is especially important for polymers, where high-frequency vibrational modes—dominated by C–H motions—impose stringent requirements for stability. To assess the impact of Δt , we developed a computational framework for T_g simulation using polyethylene (PE), chosen for its simplicity and well-characterized behavior. An amorphous single-chain 33-mer PE structure (composed of 206 atoms) was generated at 300 K using the Polymer Structure Predictor (PSP)²² and subsequently equilibrated at 600 K for 200 ps using AIMD with OTF-MLFF. Density profiles were extracted for Δt values of 0.3, 0.4, and 0.5 fs. As shown in [Figure 2a](#). Detailed information on Δt , the number of AIMD-assisted OTF-MLFF steps, configurations sampled, and CPU time has been mentioned in [Figure 2b](#). The 0.5 fs simulation exhibited large density fluctuations, indicating numerical instability, whereas the 0.3 and 0.4 fs simulations produced nearly identical, well-behaved density trajectories. Both yielded comparable sampling quality (1000 vs 1002 configurations) used to create MLFF, but with a notable difference in cost: the 0.3 fs simulation required ~30% more wall time (96 h) than the 0.4 fs run (75 h). Balancing accuracy and efficiency, a time step of $\Delta t = 0.4$ fs was identified as the optimal choice and was therefore used for all subsequent simulations in this study.

The local atomic environments sampled during the high-temperature equilibration were used to construct a production-ready MLFF. This MLFF was then used to calculate the T_g of the 33-mer amorphous PE structure obtained at the end of the 600 K AIMD with OTF-MLFF simulation. The system was cooled from 600 to 50 K in 25 K decrements, equilibrating for 100 ps at each temperature. The final configuration from each step served as the starting point for the next, and the average volume was computed from the last 50 ps of each trajectory segment to ensure thermal equilibration. The resulting volume–temperature curve ([Figure 2c](#)) was fitted to extract the glass transition temperature, with the T_g value indicated by the dotted vertical green line. The emergence of two distinct linear regimes in the volume–temperature relationship further demonstrates the model's ability to faithfully reproduce the

second-order phase transition behavior characteristic of polymer glass transitions. A 66.67% confidence interval, obtained via error-propagation analysis ([Supporting Information](#)), is shown as the orange-shaded region, while the gray-shaded regions represent the standard deviation of the volume at each temperature. The computed T_g of 281 ± 51.7 K is in reasonable agreement with the experimental range reported for a variety of polyethylene flavors^{39,40} indicated by the blue-shaded region. Fluctuations in the average volume at elevated temperatures are primarily attributed to finite-size effects inherent to the 33-mer model. Nonetheless, the MLFF exhibits strong stability and transferability across the full temperature range.

Testing Scalability across System Sizes

Next, we consider PE systems with longer chain lengths and significantly larger simulation cells containing multiple repeat units. The monomer definitions and structural details of the various amorphous PE models are provided in [SI-Table S1](#), and the corresponding T_g -derived volume–temperature curves are shown in [Figure 2c–f](#). As the system size increases, the volume/density fluctuations during cooling diminish noticeably, yielding smoother and more stable volume/density–temperature profiles. This effect is particularly apparent in the 5562-atom system ([Figure 2f](#)), where local structural variations are averaged out more effectively, leading to a more uniform macroscopic response. Moreover, as shown in [Figure 2c–f](#), the densities across different PE system sizes remain consistent, strengthening the MLFF transferability claim.

These results demonstrate that an MLFF trained on a relatively small configuration set remains robust when transferred to substantially larger systems of the same chemical composition. The model preserves accuracy and stability across both system size and temperature, underscoring the inherent scalability of the OTF-MLFF framework. Furthermore, from the T_g simulations, we extracted the coefficient of thermal expansion (CTE) in the glassy and rubbery regimes. Across all PE systems studied, the CTE ranges from 60.5×10^{-6} – 113.4×10^{-6} K⁻¹ below T_g and 438.6×10^{-6} – 671.2×10^{-6} K⁻¹ above T_g , in good agreement with the reported experimental values in the range $\sim 93 \times 10^{-6}$ K⁻¹– 600×10^{-6} K⁻¹^{40–42} (for details, refer to [SI-Table S1](#)). This confirms that the MLFF not only reproduces T_g behavior but also captures the fundamental thermophysical physics associated with polymer relaxation and thermal expansion.

We performed an independent replica T_g simulation for the 5562-atom PE system using the MLFF trained on the 206-atom data set to further assess the robustness and transferability of the developed potential. The equilibration approach employed in the replica simulation was identical to that used in the original calculation (see [Figure 2f](#)), ensuring a consistent basis for comparison. The resulting temperature-dependent volume/density profiles, along with their corresponding standard deviations, are presented in [SI-Figure S1a](#). The close agreement among the volume evolution, density trends, and fitted T_g values across the independent simulations demonstrates excellent reproducibility of the predicted glass transition behavior. Furthermore, the relatively small variations observed between the replica runs indicate that the calculated T_g is not sensitive to the choice of initial configuration or stochastic fluctuations during equilibration. These results provide strong evidence that the MLFF trained using a relatively small 206-atom reference system can reliably capture

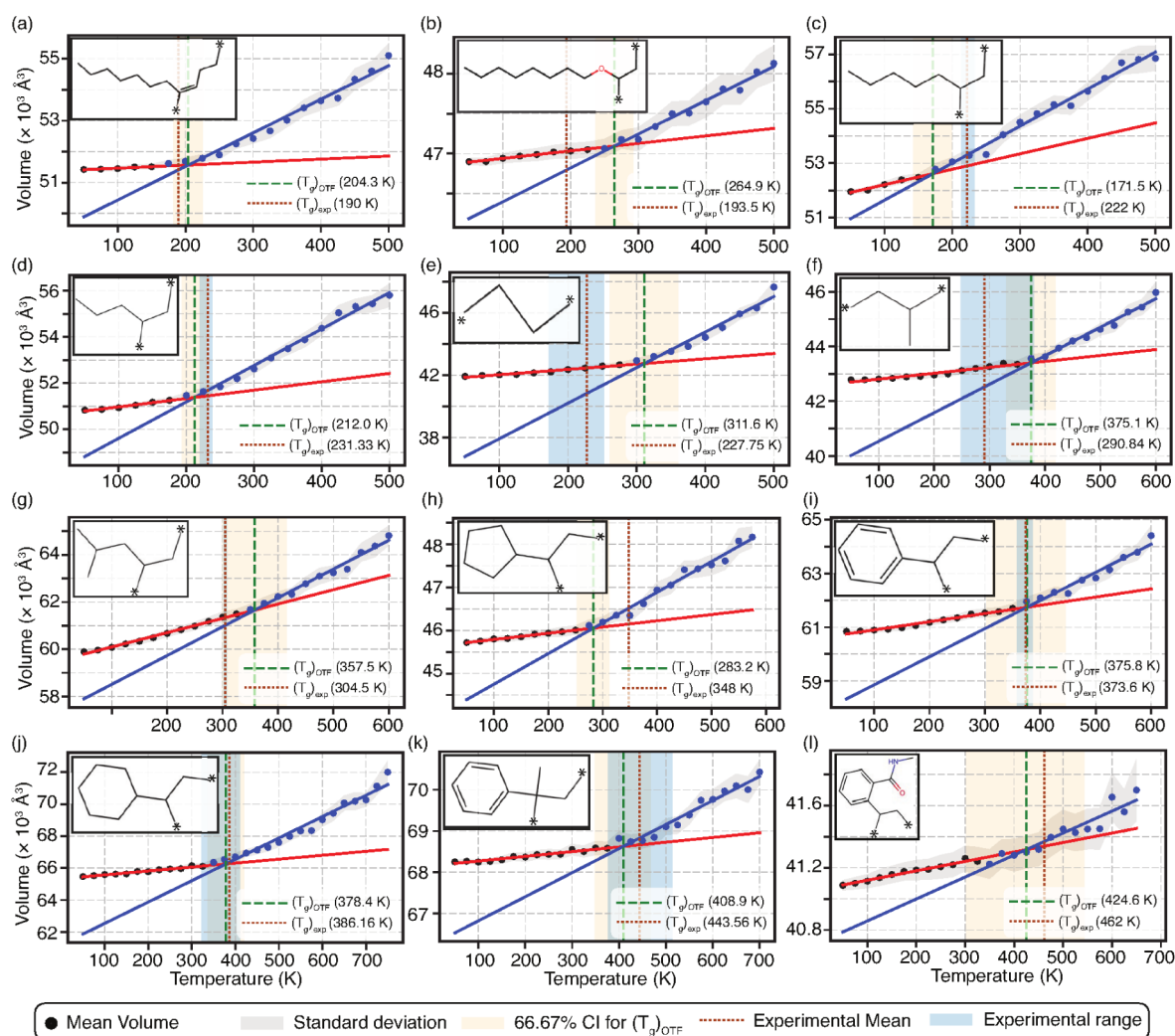


Figure 3. Volume–temperature behavior of the polymers considered in this study. Panels (a)–(l) show the volume–temperature curves for the polymers, with the corresponding chemical structure displayed in each panel. Red and blue markers with shaded gray regions denote the average volumes and standard deviations obtained from the final 50 ps of the production trajectory. Dashed green lines with shaded orange bands indicate the calculated glass-transition temperatures (T_g) and their 66.67% confidence intervals. Experimental average T_g values are indicated by red dotted lines, and the shaded blue region indicates the experimental range of the reported T_g values for comparison.

the thermodynamic behavior of substantially larger PE systems. Overall, the consistency of the replica simulations confirms the robustness, stability, and transferability of the developed MLFF for predicting glass-transition-related properties in PE.

The cooling rate is a critical parameter in T_g simulations because glass transition is inherently a kinetic phenomenon. It defines the rate at which the temperature of a material decreases with time and directly influences the extent of molecular relaxation during cooling. For a given polymer, a higher cooling rate provides less time for the polymer chains to relax toward equilibrium, generally resulting in a higher observed T_g . Consequently, experimental measurements typically employ relatively slow cooling rates (approximately 5–20 K/min) to allow sufficient molecular relaxation and structural rearrangement.^{43,44}

In contrast, atomistic molecular dynamics simulations are restricted to much higher cooling rates, typically in the range of 10^{11} – 10^{13} K/min, owing to computational limitations.⁴⁵ Despite this discrepancy of nearly 10^{12} K/min compared to experiments, previous studies have shown that simulated T_g values obtained at such cooling rates remain in reasonable

quantitative agreement with experimental measurements.^{45,46}

To evaluate the effect of cooling rate within the present OTF-MLFF framework, we performed T_g simulations for PE using two different cooling rates, 1.5×10^{13} K/min and 7.5×10^{12} K/min, comparable to those employed by Soldera et al.⁴⁶ The resulting T_g values were nearly identical, as shown in SI-Figure S1b, indicating that either cooling rate can be reliably used in subsequent OTF-MLFF simulations.

In principle, the determination of polymer T_g should be performed in the infinite-chain-length limit. However, atomistic simulations are restricted to finite chain lengths because of computational cost. To assess the adequacy of the chosen system size, we conducted a series of T_g simulations for PE using the developed MLFF, considering 27 polymer chains with chain lengths ranging from 5 to 40 monomer units (5-, 10-, 15-, 20-, 25-, 30-, and 40-mers). The 35-mer system was omitted because results were already available for a comparable 33-mer system. The simulated T_g values were analyzed using the well-known Flory–Fox relation:⁴⁷

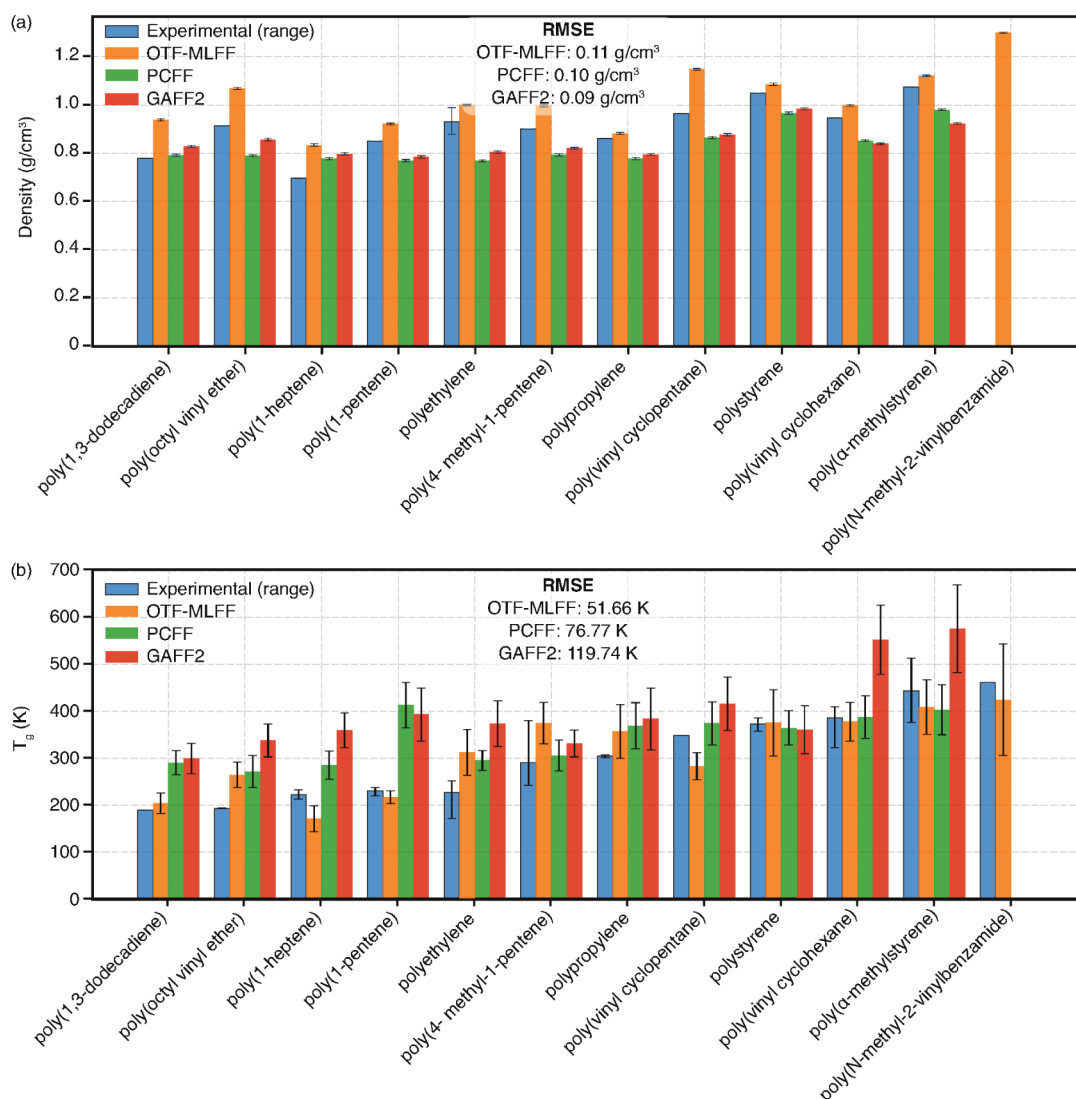


Figure 4. Comparison of OTF-MLFF approach predictions with reference data. (a) Densities at 300 K and (b) glass-transition temperatures (T_g) for the polymers investigated, benchmarked against experimental measurements and classical force-field results (PCFF and GAFF2). For the last polymer (poly(*N*-methyl-2-vinylbenzamide)), PCFF and GAFF2 simulations were not possible due to the unavailability of classical force-field parameters.

$$T_g = T_g^\infty - \frac{K}{M_n} \quad (1)$$

where T_g^∞ is the glass transition temperature at infinite molecular weight, (K) is an empirical constant, and (M_n) is the number-average molecular weight. Extrapolation of the simulation data using eq 1 yielded an infinite-chain-length glass transition temperature of 340.48 ± 16.67 K (see SI-Table S2 and Supporting Information for details). This value is in close agreement with the T_g obtained for the system containing 27 chains and 5562 atoms (Figure 2f). The agreement suggests that a simulation cell containing approximately 5000 atoms is sufficient to capture the glass transition behavior of PE within the OTF-MLFF framework, while maintaining computational efficiency.

Validating Approach over Different Chemistries

To evaluate the generalizability of the OTF-MLFF framework across diverse chemistries, we selected a set of polymers spanning a wide range of structural complexity and

experimental T_g values. The SMILES representations⁴⁸ and molecular structures of all polymers are provided in SI-Table S2. For each system, an initial amorphous cell containing roughly 200–250 atoms was constructed using PSP; the corresponding number of monomer units and total atoms per cell are also summarized in SI-Table S2. This training system size reflects the practical computational limits of AIMD (based on the Figure 2b analysis), ensuring that each MLFF could be generated within a feasible wall time.

Each polymer was equilibrated for 200 ps at a designated training temperature (T_{training}), chosen to sample a broad and representative range of atomic configurations suitable for subsequent T_g simulations. The number of configurations collected during training (N_{sampled}) varied by chemical complexity and is listed in SI-Table S2. Leveraging the insights drawn from the PE case, we then considered significantly large structures for temperature-ramped simulations, typically 4600–8000 atoms constructed by replicating and combining multiple chains of each polymer.

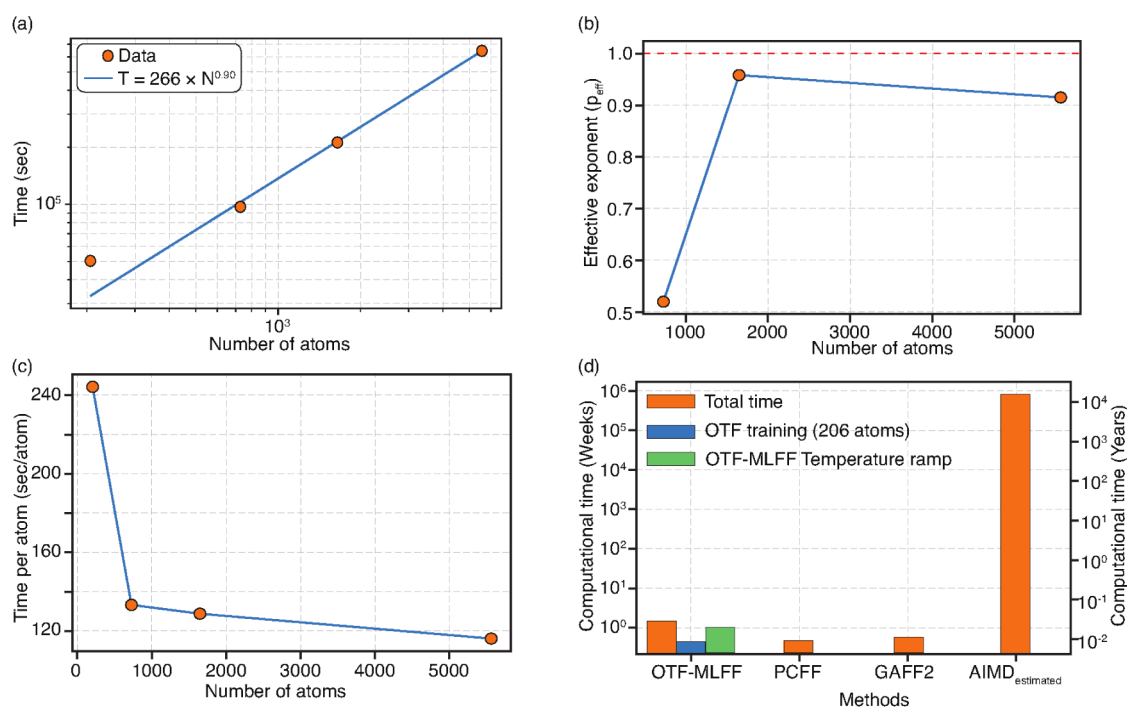


Figure 5. Scaling behavior and computational benchmarking of the simulations. (a) log–log plot of simulation time with system size (N), (b) effective exponent (p_{eff}) as a function of N , (c) simulation time per atom versus N , and (d) comparison of total simulation times for OTF-MLFF, the Polymer Consistent Force Field (PCFF), the General AMBER Force Field (GAFF2), and ab initio molecular dynamics (AIMD).

The resulting volume–temperature curves, shown in Figure 3a–l, include the polymer structures, the standard deviation in equilibrated volumes (gray-shaded regions), and the fitted T_g values with their associated confidence intervals (vertical dashed green lines). Corresponding density–temperature plots for all cases are provided in SI-Figure S3. Across all 12 polymers studied, the predicted T_g values agree closely with the experimental values and lie within or close to the respective confidence bounds. This strong correlation underscores the predictive reliability of the OTF-MLFF approach, which successfully captures the glass transition behavior of polymers with aromatic, aliphatic, branched, and heteroatom-containing backbones. Because the entire workflow is anchored in first-principles calculations and involves no empirical fitting or prior knowledge of the experimental T_g , the method is inherently transferable and broadly applicable. It provides a robust and universal pathway for simulating thermophysical properties of chemically diverse polymer systems well beyond the reach of traditional force-field approaches.

The physical and chemical fidelity of the developed MLFFs is further confirmed by the excellent agreement between the predicted and experimental densities and glass transition temperatures across all polymers studied. For representative systems such as PE and polystyrene (PS), the experimental densities at 300 K of 0.88–0.93 and 1.05 g/cm³, respectively^{49,50} come close to the MLFF estimates of 0.99 and 1.09 g/cm³, respectively. Similar levels of consistency are observed for the remaining polymers, as illustrated in Figure 4a and b, which compare the experimentally measured and MLFF-predicted T_g values and 300 K densities. Further, the thermal expansion information was extracted from the volume–temperature plots (Figure 3) of the respective polymers and is reported in SI-Table S3. These results highlight that the OTF-MLFF workflow effectively captures the essential chemical interactions and thermophysical behavior even in systems with

complex chemistries, long chain lengths, and multiple repeat units, thereby reinforcing its scalability, robustness, and versatility.

To further benchmark performance, we compared OTF-MLFF predictions with classical force-field simulations using PCFF and GAFF2. The RMSE values for density relative to experiment at room temperature are comparable across methods: 0.11 g/cm³ (OTF-MLFF), 0.10 g/cm³ (PCFF), and 0.09 g/cm³ (GAFF2), indicating that all three approaches reproduce equilibrium densities reasonably well at room temperature (Figure 4). In contrast, the advantage of OTF-MLFF becomes pronounced for glass-transition prediction, which is a straight test of performance over a broad temperature range: the RMSE for T_g is significantly lower at 51.66 K for OTF-MLFF, compared to 76.77 K for PCFF and 119.74 K for GAFF2. The OTF-MLFF performance also surpasses refined OPLS⁵¹ and OPLS3e⁵² force fields, which report T_g RMSE values of 98.5 K⁵¹. It is important to note that the reported RMSE in the predicted glass transition temperatures does not arise solely from the OTF-MLFF. It also reflects uncertainties inherited from the underlying DFT reference calculations, including the choice of exchange–correlation functional, which can introduce systematic errors in describing the intermolecular interactions that influence polymer glass transition behavior.

Importantly, classical force-field simulations could not be performed for poly(*N*-methyl-2-vinylbenzamide) due to missing parameters, whereas OTF-MLFF successfully handled this system without modification of the simulation approach, reflecting its intrinsic ab initio-derived generality. Collectively, these comparisons underscore the ability of OTF-MLFFs to accurately capture polymer thermophysical properties even for chemistries that lie beyond the domain of traditional force fields or even stand-alone MLFFs. A useful aspect of our approach is that the second-order phase transition can be

identified directly from the volume–temperature curves, without relying solely on mathematical fitting. This offers an alternative perspective to the recently introduced Vivace MLFF method,¹⁰ particularly in cases where the transition is difficult to detect. In their work, four of the ten polymers studied (PODPM, PVMS, PAN, and PTPC) exhibit nearly linear volume–temperature behavior in the entire temperature range considered with no discernible change in slope, suggesting that a glass transition may not be present according to their simulations. From a physical standpoint, the absence of any perceptible change in thermal expansion would indicate that fitting for an expected transition may not yield in reliable transition temperature. It is a bit surprising that Vivace approach does not produce a discernible transition, even though they bracket the simulated temperature range around experimental T_g value for the system studied. In our workflow, careful sampling from diverse regions of phase space during MLFF training facilitates the unambiguous detection of glassy and rubbery regimes when such transitions exist, even though an unbiased range of temperatures is used to capture the glass transition. This allows the method to avoid assigning T_g where no transition is apparent and to provide more reliable estimates when a transition is observable.

Computational Time Advantage

A notable outcome of this work is the ability of MLFFs trained on only ~1000 AIMD configurations per polymer chemistry from 200–250-atom cells to remain accurate when applied to much larger systems. Using PE as a representative example, we now examined how the computational cost evolves with system size (Figure 5a). These MLFFs exhibit a near-linear (almost sublinear) scaling of $N^{0.90}$ (N is the total number of atoms), in sharp contrast to the N^2 – N^3 scaling of conventional DFT. This behavior highlights the practical scalability of the approach and its suitability for simulations of polymer structures containing several thousand atoms.

In order to confirm the (sub)linear scaling of the approach, the effective exponent (p_{eff}) and time per atom (T/N) as a function of number of atoms in the simulation (N) were calculated. p_{eff} is extracted from two consecutive points for the two neighboring system sizes N_i and N_{i+1} with corresponding computational times T_i and T_{i+1} using the following eq 2:

$$p_{\text{eff}}(N_i) = \frac{\log(T_{i+1}/T_i)}{\log(N_{i+1}/N_i)} \quad (2)$$

Figure 5b further examines the effective scaling exponent (p_{eff}) across system sizes. In all cases, p_{eff} remains below or close to unity, confirming sub (linear) behavior. The initial rise, followed by a gradual potential saturation close to unity, reflects the crossover from small-system overheads to the regime where the MLFF achieves its characteristic efficient scaling. A complementary assessment using the simulation time per atom (Figure 5c) similarly shows a monotonic decrease with increasing N . This behavior underscores both the scalability and the cost-effectiveness of the method while preserving *ab initio* accuracy for large-scale polymer simulations.

The efficiency of the approach becomes particularly evident when comparing the computational cost of T_g simulations across methods. For polyethylene, training the OTF-MLFF using a 206-atom system for 200 ps and subsequently applying it to equilibrate a 5562-atom supercell for 100 ps at 23 temperature points (from 600 to 50 K in 25 K intervals)

amounts to roughly 6.25 million simulation time-steps. Using the OTF-MLFF framework, these steps were completed in approximately 10 days on 32 cores of an Intel(R) Xeon(R) Gold 6140 CPU @ 2.30 GHz. The distribution of wall time between MLFF training and the T_g cooling trajectory is shown in Figure 5d.

In stark contrast, performing the same number of simulation steps for the 5562-atom PE structure using AIMD would require an estimated ~15,000 years on the same hardware. This estimate was determined by simply multiplying the time taken for a single time step, in the same hardware for the same system size, by 6.25 million. Given that T_g simulations inherently rely on serial cooling trajectories, such AIMD-based calculations become computationally prohibitive and likely cannot be practically scaled even on modern high-performance computing architectures. With the hybrid OTF-MLFF approach, we achieve an extraordinary speedup of approximately 10^6 , while preserving *ab initio* fidelity in energies, forces, and stresses. Importantly, this acceleration does not compromise the predictive reliability of the model, underscoring the real-world applicability of OTF-MLFF for large-scale polymer simulations and the accurate prediction of complex thermophysical properties.

Opportunities

While the OTF-MLFF framework is highly promising, several avenues remain for further refinement. One important direction is enhancing force-field robustness by integrating training data sampled across a broader temperature range, which could further improve transferability and performance in regimes with significant structural or dynamical variation. A temperature-diverse training strategy may reduce volume fluctuations during cooling trajectories, thereby improving the precision of T_g estimation. Although the adaptive OTF-MLFF framework continuously expands the training set by incorporating newly encountered out-of-distribution (OOD) configurations through on-the-fly DFT calculations, previously unseen OOD configurations may still arise during sufficiently long simulations. Developing more robust OOD detection and uncertainty quantification strategies, therefore, represents an important direction for future research and could further improve the reliability and transferability of OTF-MLFF for long-time scale simulations. However, achieving this within the current VASP-based OTF workflow is computationally demanding, and the benefits relative to the required cost are not yet fully established. This challenge highlights a deeper need for algorithmic and code-level optimization, both in the underlying AIMD engine and in the MLFF construction pipeline, to reduce overheads and enable more extensive sampling. Another direction and challenge is to develop a transferable MLFF that can be created by combining previously generated OTF-MLFFs without performing new simulations. Because a model for each polymer is trained independently, differences in descriptor distributions, energy baselines, and configuration-space coverage complicate direct integration. Several computational strategies can address this limitation, including descriptor harmonization through normalization or alignment, posthoc energy, force, and stress rereferencing to ensure consistency, and data-fusion approaches that merge and prune training sets to produce a chemically diverse yet efficient reference database. Alternatively, modular or mixture-of-experts architectures could blend multiple MLFFs based on local environments, enabling

smooth interpolation across polymer chemistries. Collectively, these approaches offer a pathway toward constructing a broadly transferable, low-cost MLFF by reusing existing OTF-generated data without repeated AIMD sampling.

Looking ahead, the efficiency and near-first-principles accuracy of OTF-MLFFs open the door to entirely new classes of polymer simulations that were previously inaccessible. These include long-term thermal protocols, such as repeated cooling–heating cycles for accurate T_g prediction, as well as mechanical deformation and fracture simulations that capture bond breaking and strain localization. Additionally, molecular and ionic transport studies may be conceived that involve the diffusion of gases, solvents, and ions through dense or heterogeneous polymer matrices. OTF-MLFFs may also enable simulations of degradation mechanisms and chemical transformations. By supporting these computationally demanding studies at scale, while maintaining near-quantum-level fidelity, OTF-MLFFs have the potential to significantly broaden the scope of polymer modeling and provide quantitative, mechanistic insights into phenomena that have long been beyond reach.

CONCLUSION

In conclusion, we present an on-the-fly machine-learned force field (OTF-MLFF) computational framework for predicting the glass transition temperature (T_g) of polymers from atomistic simulations. By combining first-principles calculations, active learning, and MLFFs, the approach enables efficient simulations at scales that are challenging for direct ab initio methods while requiring only a limited number of training configurations. Across the polymers examined in this study, the resulting force fields reproduced physically reasonable density–temperature behavior and enabled direct estimation of T_g from continuous cooling simulations without initialization near experimental values. These results demonstrate the potential of OTF-MLFFs as a practical and data-efficient framework for studying thermophysical behavior in polymers and provide a foundation for future applications to broader polymer chemistries and properties.

ASSOCIATED CONTENT

Data Availability Statement

First-principles calculations, configuration sampling, and the generation of on-the-fly machine learning force fields (OTF-MLFFs) were performed using the commercially available Vienna Ab initio Simulation Package (VASP). The complete set of molecular structures and atomic configurations used for force-field training, validation, and glass transition temperature (T_g) simulations will be made publicly available on GitHub <https://github.com/Ramprasad-Group/OTF-MLFF-Tg> repository upon publication of this work to ensure reproducibility. These structures served as input geometries for VASP calculations, with the computational parameters and electronic-structure settings described in the Computational Methodology section applied consistently throughout the study. The resulting simulation data were subsequently analyzed to determine the glass transition temperature, density, and coefficient of thermal expansion using the postprocessing workflow described in the Supporting Information. The Supporting Information additionally provides a detailed description of the data analysis procedures, fitting methodologies, and criteria used to extract these thermophysical

properties from the molecular dynamics trajectories, enabling independent reproduction of the reported results.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.6c02116>.

(1) Monomer units, polymer chains, number of atoms, corresponding glass transition temperature (T_g), and coefficient of thermal expansion of various PE structures (SI-Table S1), (2) details of monomer units, number of atoms in the training unit cell, number of configurations sampled during MLFF training, training temperatures, and simulated larger structure used for T_g calculations of the polymers (SI-Table S2) studied in this work, (3) replica and two different cooling rate T_g simulations for PE, (4) weighted Flory–Fox extrapolation of T_g to infinite molecular weight for PE, (5) step-by-step MLFF generation approach and extracting glass transition temperature, and (6) OTF-MLFF estimated coefficient of thermal expansion values below and above glass transition (SI-Table S3). A description of error propagation estimation and a density–temperature plot for the simulated polymers (SI-Figure S3) (PDF)

AUTHOR INFORMATION

Corresponding Author

Rampi Ramprasad – School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States; orcid.org/0000-0003-4630-1565; Email: ramprasad@gatech.edu

Authors

Ashtosh Srivastava – School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States; orcid.org/0000-0002-7489-4065

Sakshi Agarwal – School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States

Shivank Shukla – School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States

Harikrishna Sahu – School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States; orcid.org/0000-0001-5458-9488

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jcim.6c02116>

Author Contributions

A.S. designed and established the overall simulation framework, performed the DFT and OTF-MLFF calculations, carried out data analysis, and prepared, revised and finalized the manuscript. S.A. contributed to the data analysis and assisted in preparing the original draft. S.S. performed the classical force-field simulations. H.S. supported the generation and preparation of polymer structures. R.R. conceived and supervised the project, provided overall guidance, and contributed to the interpretation of results.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is financially supported by the Office of Naval Research through a multidisciplinary university research initiative (MURI) grant N00014-20-1-2586, and the National Science Foundation through an NSF-SPEED grant 2515411. Computations were performed at Expanse (San Diego Supercomputing Center) through an allocation (DMR080044) from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program. Discussions with Ayush Jain and Anagha Savit are also gratefully acknowledged.

REFERENCES

- (1) Tran, H.; Gurnani, R.; Kim, C.; Pilania, G.; Kwon, H.-K.; Lively, R. P.; Ramprasad, R. Design of functional and sustainable polymers assisted by artificial intelligence. *Nat. Rev. Mater.* **2024**, *9*, 866–886.
- (2) Muench, S.; Wild, A.; Friebe, C.; Häupler, B.; Janoschka, T.; Schubert, U. S. Polymer-Based Organic Batteries. *Chem. Rev.* **2016**, *116*, 9438–9484.
- (3) Hager, M. D.; Bode, S.; Weber, C.; Schubert, U. S. Shape memory polymers: Past, present and future developments. *Prog. Polym. Sci.* **2015**, *49–50*, 3–33.
- (4) Huan, T. D.; Boggs, S.; Teyssedre, G.; Laurent, C.; Cakmak, M.; Kumar, S.; Ramprasad, R. Advanced polymeric dielectrics for high energy density applications. *Prog. Mater. Sci.* **2016**, *83*, 236–269.
- (5) Pan, M.; Pan, C.; Li, C.; Zhao, J. A review of membranes in proton exchange membrane fuel cells: Transport phenomena, performance and durability. *Renewable Sustainable Energy Rev.* **2021**, *141*, 110771.
- (6) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, 2453.
- (7) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **2023**, *14*, 579.
- (8) Kovács, D. P.; Moore, J. H.; Browning, N. J.; Batatia, I.; Horton, J. T.; Pu, Y.; Kapil, V.; Witt, W. C.; Magdau, I.-B.; Cole, D. J.; Csányi, G. MACE-OFF: Short-Range Transferable Machine Learning Force Fields for Organic Molecules. *J. Am. Chem. Soc.* **2025**, *147*, 17598–17611.
- (9) Qi, Y.; Gong, W.; Yan, Q. Bridging deep learning force fields and electronic structures with a physics-informed approach. *npj Comput. Mater.* **2025**, *11*, 177.
- (10) Simm, G. N. C.; Hélie, J.; Schulz, H.; Chen, Y.; Simeon, G.; Kuzina, A.; Martinez-Baez, E.; Gasparotto, P.; Tocci, G.; Chen, C. et al. SimPoly: Simulation of Polymers with Machine Learning Force Fields Derived from First Principles. *arXiv* **2025**.
- (11) Botu, V.; Ramprasad, R. Adaptive machine learning framework to accelerate ab initio molecular dynamics. *Int. J. Quantum Chem.* **2015**, *115*, 1074–1083.
- (12) Jinnouchi, R.; Karsai, F.; Kresse, G. On-the-fly machine learning force field generation: Application to melting points. *Phys. Rev. B* **2019**, *100*, 014105.
- (13) Roos, Y. H. Glass Transition Temperature and Its Relevance in Food Processing. *Annu. Rev. Food Sci. Technol.* **2010**, *1*, 469–496.
- (14) Han, J.; Gee, R. H.; Boyd, R. H. Glass transition temperatures of polymers from molecular dynamics simulations. *Macromolecules* **1994**, *27*, 7781–7784.
- (15) Park, J.-M.; Park, C. S.; Kwak, S. K.; Sun, J.-Y. Glass transition temperature as a unified parameter to design self-healable elastomers. *Sci. Adv.* **2024**, *10*, No. eadp0729.
- (16) Yu, K.-Q.; Li, Z.-S.; Sun, J. Polymer structures and glass transition: A molecular dynamics simulation study. *Macromol. Theory Simul.* **2001**, *10*, 624–633.
- (17) Buchholz, J.; Paul, W.; Varnik, F.; Binder, K. Cooling rate dependence of the glass transition temperature of polymer melts: Molecular dynamics study. *J. Chem. Phys.* **2002**, *117*, 7364–7372.
- (18) Mohammadi, M.; Fazli, H.; Karevan, M.; Davoodi, J. The glass transition temperature of PMMA: A molecular dynamics study and comparison of various determination methods. *Eur. Polym. J.* **2017**, *91*, 121–133.
- (19) Vandermause, J.; Torrisi, S. B.; Batzner, S.; Xie, Y.; Sun, L.; Kolpak, A. M.; Kozinsky, B. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *npj Comput. Mater.* **2020**, *6*, 20.
- (20) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169.
- (21) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.
- (22) Sahu, H.; Shen, K.-H.; Montoya, J. H.; Tran, H.; Ramprasad, R. Polymer Structure Predictor (PSP): A Python Toolkit for Predicting Atomic-Level Structural Models for a Range of Polymer Geometries. *J. Chem. Theory Comput.* **2022**, *18*, 2737–2748.
- (23) Landrum, G. RDKit: Open-Source Cheminformatics Software. **2024**; <https://www.rdkit.org>.
- (24) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.
- (25) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
- (26) Abbott, L. J.; Hart, K. E.; Colina, C. M. Polymatic: a generalized simulated polymerization algorithm for amorphous polymers. *Theor. Chem. Acc.* **2013**, *132*, 1334.
- (27) Jinnouchi, R.; Lahnsteiner, J.; Karsai, F.; Kresse, G.; Bokdam, M. Phase Transitions of Hybrid Perovskites Simulated by Machine-Learning Force Fields Trained on the Fly with Bayesian Inference. *Phys. Rev. Lett.* **2019**, *122*, 225701.
- (28) Jinnouchi, R.; Karsai, F.; Verdi, C.; Asahi, R.; Kresse, G. Descriptors representing two- and three-body atomic distributions and their effects on the accuracy of machine-learned inter-atomic potentials. *J. Chem. Phys.* **2020**, *152*, 234102.
- (29) Hoover, W. G.; Ladd, A. J. C.; Moran, B. High-Strain-Rate Plastic Flow Studied via Nonequilibrium Molecular Dynamics. *Phys. Rev. Lett.* **1982**, *48*, 1818–1820.
- (30) Evans, D. J. Computer “experiment” for nonlinear thermodynamics of Couette flow. *J. Chem. Phys.* **1983**, *78*, 3297–3302.
- (31) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45*, 1196–1199.
- (32) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (33) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (34) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (35) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50*, 17953.
- (36) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758.
- (37) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.
- (38) Pulay, P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.* **1980**, *73*, 393–398.

- (39) Fakirov, S.; Krasteva, B. On the glass transition temperature of polyethylene as revealed by microhardness measurements. *J. Macromol. Sci., Part B: Phys* **2000**, *39*, 297–301.
- (40) Ohlberg, S. M.; Fenstermaker, S. S. The determination of the glass transition temperature of polyethylene by x-ray diffraction. *J. Polym. Sci* **1958**, *32*, 514–516.
- (41) Passive-Components.eu Coefficient of Linear Thermal Expansion on Polymers Explained. <https://passive-components.eu/coefficient-of-linear-thermal-expansion-on-polymers-explained/>. Accessed 29 November 2025.
- (42) Material Property Tables: PE (LDPE/HDPE). https://plastecprofiles.com/uploads/1/3/0/4/13043900/material_property_tables_pe.pdf, Accessed 29 November 2025.
- (43) Gaisford, S.; Kett, V.; Haines, P. *Principles of thermal analysis and calorimetry*. Royal society of chemistry, 2016.
- (44) Menczel, J. D.; Prime, R. B. *Thermal Analysis of Polymers: Fundamentals and Applications*. Wiley 2009
- (45) Wang, Y.-C.; Zhang, J.-F.; Chiu, M.-H.; Li, J.-H.; Jui, C.-Y.; Yang, T.-H.; Lee, W.-J. Molecular-weight and cooling-rate dependence of polymer thermodynamics in molecular dynamics simulation. *Polym. J* **2021**, *53*, 455–462.
- (46) Soldera, A.; Metatla, N. Glass transition of polymers: Atomistic simulation versus experiments. *Phys. Rev. E* **2006**, *74*, 061803.
- (47) Fox, T. G.; Flory, P. J. Second-Order Transition Temperatures and Related Properties of Polystyrene. I. Influence of Molecular Weight. *J. Appl. Phys* **1950**, *21*, 581–591.
- (48) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci* **1988**, *28*, 31–36.
- (49) Cheremisinoff, N. P. *Condensed Encyclopedia of Polymer Engineering Terms*. Elsevier, 2001; pp. 200–255.
- (50) Gehlsen, M. D.; Weimann, P. A.; Bates, F. S.; Harville, S.; Mays, J. W.; Wignall, G. D. Synthesis and characterization of poly-(vinylcyclohexane) derivatives. *J Polym Sci B Polym Phys* **1995**, *33*, 1527–1536.
- (51) Afzal, M. A. F.; Browning, A. R.; Goldberg, A.; Halls, M. D.; Gavartin, J. L.; Morisato, T.; Hughes, T. F.; Giesen, D. J.; Goose, J. E. High-Throughput Molecular Dynamics Simulations and Validation of Thermophysical Properties of Polymers for Various Applications. *ACS Appl. Polym. Mater* **2021**, *3*, 620–630.
- (52) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput* **2019**, *15*, 1863–1874.