

Bridging machine learning and glassy dynamics theory for predictive polymer modeling

Cite as: J. Appl. Phys. 138, 044703 (2025); doi: 10.1063/5.0280443

Submitted: 12 May 2025 · Accepted: 8 July 2025 ·

Published Online: 25 July 2025



View Online



Export Citation



CrossMark

Anh D. Phan,^{1,2,a)} Ngo T. Que,² Nguyen T. T. Duyen,¹ Phan Thanh Viet,³ Quach K. Quang,³ and Baicheng Mei⁴

AFFILIATIONS

¹Faculty of Materials Science and Engineering, Phenikaa University, Hanoi 12116, Vietnam

²Phenikaa Institute for Advanced Study, Phenikaa University, Hanoi 12116, Vietnam

³Division of Physics, School of Education, Dong Thap University, Cao Lanh 870000, Vietnam

⁴School of Materials Science and Engineering, Beijing Institute of Technology, Beijing 100081, China

Note: This paper is part of the Special Topic on Integrating Data Science and Computational Materials Science.

a) Author to whom correspondence should be addressed: anh.phanduc@phenikaa-uni.edu.vn

ABSTRACT

Understanding and predicting the glassy dynamics of polymers remain fundamental challenges in soft matter physics. While the elastically collective nonlinear Langevin equation theory has been successful in describing relaxation dynamics, its practical application to polymers depends on thermal mapping, to connect the theory with the experiment, which, in turn, requires detailed thermodynamic data. Such data may not be available for chemically complex or newly designed polymers. In this work, we propose a simple approach that integrates machine learning-predicted glass transition temperatures (T_g) with a simplified thermal mapping based on an effective thermal expansion coefficient to overcome these limitations. This approach can provide quantitatively accurate predictions of relaxation dynamics across a broad range of polymers. Rather than replacing the original thermal mapping, our method complements it by trading formal rigor for computational efficiency and broader applicability in high-throughput screening and materials with limited available data. Moreover, we introduce a physically motivated modification to the thermal mapping that resolves discrepancies in the description of low- T_g polymers. Our results establish a generalizable approach for predictive modeling of glassy polymer dynamics and point toward new directions for theory-guided materials discovery.

26 July 2025 10:38:18

© 2025 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0280443>

I. INTRODUCTION

Investigating the glassy dynamics of polymers, thermal liquids, organic materials, and amorphous drugs is critically important due to their broad applications across industrial, pharmaceutical, and technological sectors. Glassy dynamics strongly influence material stability, mechanical properties, and functional performance, particularly for amorphous pharmaceuticals where the glass transition temperature directly affects drug solubility, dissolution rates, and bioavailability.^{1–3} In polymers and organic materials, understanding structural relaxation processes and glass transition phenomena enables better control over processing conditions, mechanical durability, and thermal stability, essential for optimizing material performance in applications ranging from packaging to optoelectronics.^{4,5}

Despite extensive experimental and theoretical studies, many fundamental aspects including the microscopic mechanisms governing the alpha and beta relaxation processes remain incompletely understood, necessitating continued research to advance theoretical frameworks and predictive models.^{6,7}

To address these unresolved issues comprehensively, researchers have employed a variety of experimental and theoretical methods to investigate the activated dynamics of both supercooled liquids and glasses and the glass transition temperature. Experimentally, techniques such as differential scanning calorimetry (DSC), broadband dielectric spectroscopy (BDS), neutron scattering, and nuclear magnetic resonance have provided critical insights into relaxation behaviors, molecular mobility, and dynamic heterogeneities near the glass transition.^{2,8} Complementing these experiments, theoretical and

computational approaches such as mode coupling theory, random first-order transition theory, and the elastically collective nonlinear Langevin equation (ECNLE) theory have provided powerful frameworks for interpreting and predicting the underlying physics of glass transition and structural relaxation dynamics.^{7,9,10} Recent advances in molecular dynamics (MD) simulations^{7,11,12} and machine/deep learning^{13–23} have further enhanced our ability to unravel microscopic mechanisms and quantitatively link experimental observations with theoretical predictions. These approaches significantly deepen and accelerate our understanding of glassy systems.

Among the theoretical frameworks, the ECNLE theory has emerged as a powerful and versatile approach for describing glassy dynamics across a wide range of materials.^{7,11,12,24–43} By quantitatively capturing the interplay between local cage-scale dynamics and longer-range collective elastic effects, the ECNLE theory has successfully explained the temperature dependence of structural relaxation times, viscosity growth, diffusion constants, and dynamic shear moduli in thermal liquids,^{7,12,28–30,32} polymers,^{11,12,24–27,31–34,38,40–43} colloidal suspensions,^{27,28,32} metallic glasses,^{36,37,39} amorphous drugs,³⁵ and nanocomposites.^{36,37} Furthermore, the ECNLE theory has been extended to investigate the confinement effects in thin polymer films.^{11,12,24–27,38} The calculation of temperature-dependent relaxation times and diffusion constants within the ECNLE theory requires a thermal mapping from material density to temperature. This conversion is constructed from the experimental equation-of-state data or using the experimental glass transition temperature values.^{7,11,12,24–38} Recent developments coupling ECNLE calculations with simulation data and machine learning models have further enhanced the predictive accuracy for metallic glasses^{37,39} without the need for adjustable or empirically fitted parameters.

Despite the significant successes of the ECNLE theory^{7,11,12,24–38} in describing the glassy dynamics of various materials, several critical challenges still hinder its broad predictive applicability. A major limitation lies in the dependence of ECNLE calculations on accurate knowledge of the glass transition temperature of materials, which is often unavailable for new or unsynthesized polymers. The traditional experimental determination of T_g is time-consuming, costly, and infeasible for high-throughput material discovery. Furthermore, even when accurate T_g values are available, the current ECNLE formulation struggles to accurately predict the structural relaxation behavior of certain polymers such as polydimethylsiloxane. This strongly suggests a need for further theoretical development concerning polymer-specific physics or mechanisms.

In this work, we systematically address the aforementioned challenges by combining machine learning (ML) techniques with the ECNLE theoretical framework. First, we train and validate several ML models on a comprehensive and chemically diverse dataset to accurately predict T_g . We then benchmark the predictive performance of these models against previous studies to identify the most effective ML approach for T_g prediction. Subsequently, we integrate these ML-derived T_g values into the ECNLE theory to quantitatively predict the temperature-dependent structural relaxation dynamics of a broad class of polymers. Finally, we propose a physically motivated thermal mapping to accurately describe the temperature dependence of relaxation dynamics of polymers that were previously not well-captured. This work builds on our recent ML-ECNLE approach developed for metallic glasses and extends it

to a wide range of polymer systems. Thereby, we significantly expand the predictive scope and generality of the ECNLE theory.

II. THEORETICAL BACKGROUND

A. The ECNLE theory

Within the framework of the original ECNLE theory, the activation events in any amorphous material are theoretically investigated using a hard-sphere fluid model.^{7,11,12,24–39} Amorphous materials are modeled as a dense assembly of hard spheres characterized by the density of particle numbers, ρ , and the diameter of the particles, d . The static structure factor, $S(q)$, and the radial distribution function, $g(r)$, are calculated based on integral equation theory with the Percus-Yevick closure approximation.⁴⁴ The local dynamics of a tagged particle is primarily governed by the interactions with neighboring particles, random thermal noise, and frictional force, as described by a nonlinear stochastic equation of motion. Solving this equation yields the dynamic free energy, $F_{\text{dyn}}(r)$, which quantifies the effective caging constraints experienced by the particle^{7,11,12,24–39}

$$\frac{F_{\text{dyn}}(r)}{k_B T} = -3 \ln \left(\frac{r}{d} \right) - \int_0^{\infty} \frac{q^2 d^3 [S(q) - 1]^2}{12\pi\Phi[1 + S(q)]} \exp \left[-\frac{q^2 r^2 (1 + S(q))}{6S(q)} \right] dq, \quad (1)$$

where T is the temperature, k_B is the Boltzmann constant, r is the particle displacement, q is the wavevector, and $\Phi = \rho\pi d^3/6$ is the volume fraction. The first term on the right-hand side of Eq. (1) describes an ideal fluid-like behavior, while the second term accounts for the confinement effect due to neighboring particles (caging constraint).

The dynamic free energy provides critical understanding of the mechanisms that govern the local motion of the particles. At low volume fractions ($\Phi < 0.432$), the particles exhibit the characteristic behavior of an ideal fluid and freely diffuse with minimal resistance due to the weak interactions between particles.^{26–30} As the system becomes denser, stronger interparticle forces lead to the emergence of cage-like structures that transiently trap a central particle. The characteristic size of this cage, r_{cage} , is identified as the position of the first minimum in the radial distribution function, $g(r)$. Consequently, the local energy barrier in the dynamic free energy increases as illustrated in Fig. 1. The height of this barrier is determined as $F_B = F_{\text{dyn}}(r_B) - F_{\text{dyn}}(r_L)$, where r_L and r_B correspond to the localization length and the barrier position of $F_{\text{dyn}}(r)$, respectively. The displacement between these two dynamic positions, $\Delta r = r_B - r_L$, characterizes the jump distance of the particle required to escape the local cage.

To successfully escape its local cage, a particle must overcome not only its immediate constraints but also facilitate cooperative rearrangements of the surrounding particles to generate sufficient free volume.^{7,11,12,24–39} This collective motion produces a radially outward displacement field, $u(r)$, which originates at the cage surface and propagates into the surrounding medium. Within the framework of Lifshitz's continuum elasticity theory,⁴⁵ the displacement field for $r \geq r_{\text{cage}}$ can be analytically expressed as

26 July 2025 10:38:18

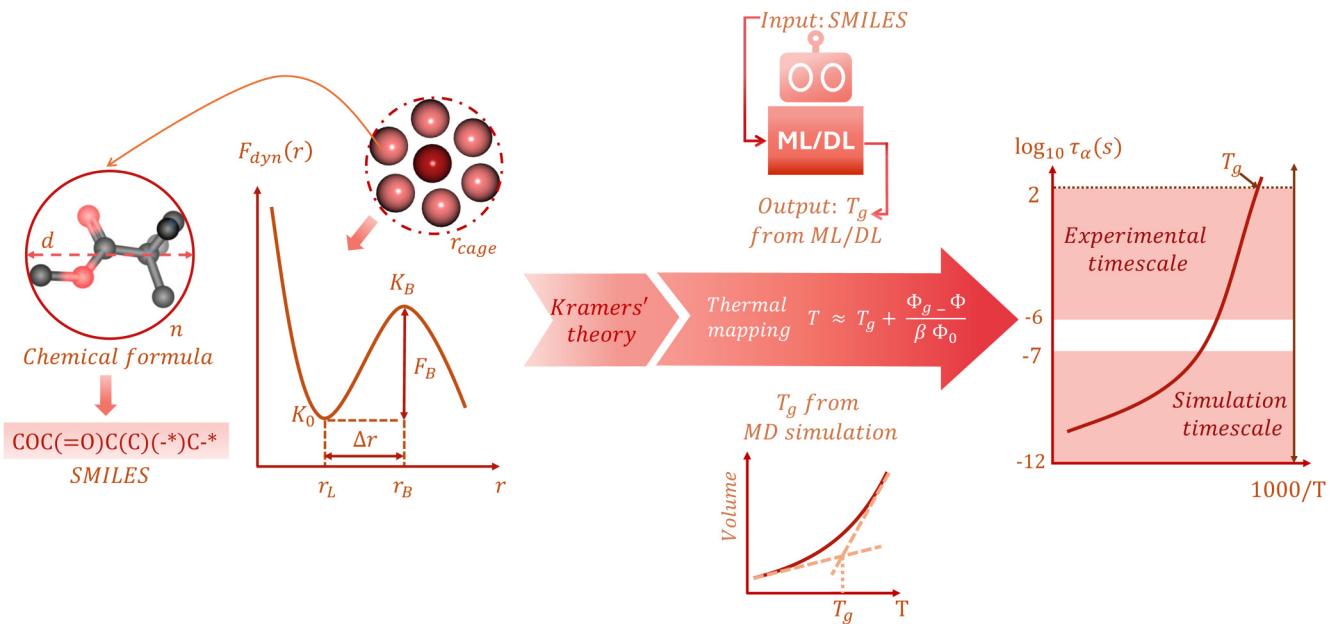


FIG. 1. Workflow for predicting the glass transition temperature of polymers using molecular dynamics simulations and deep learning/machine learning, after which the predicted T_g is used in the ECNLE theory to compute the structural relaxation time and diffusion coefficients.

$$u(r) = \frac{\Delta r_{\text{eff}} r_{\text{cage}}^2}{r^2}, \quad (2)$$

where Δr_{eff} is the effective amplitude of the displacement field. The analytical expression of Δr_{eff} is^{28–30}

$$\Delta r_{\text{eff}} = \frac{3}{r_{\text{cage}}^3} \left[\frac{r_{\text{cage}}^2 \Delta r^2}{32} - \frac{r_{\text{cage}} \Delta r^3}{192} + \frac{\Delta r^4}{3072} \right]. \quad (3)$$

Since $u(r)$ is generally small, particle displacements are treated as harmonic oscillations. The local elastic energy for a single oscillator is $K_0 u^2(r)/2$, where $K_0 = |\partial^2 F_{\text{dyn}}(r)/\partial r^2|_{r=r_L}$ is a packing-fraction-dependent spring constant. The total collective elastic barrier is obtained by integrating the elastic energy contributions from all particles located outside the cage^{7,11,12,24–39}

$$F_e = 4\pi\rho \int_{r_{\text{cage}}}^{\infty} r^2 g(r) \frac{K_0 u^2(r)}{2} dr. \quad (4)$$

The structural relaxation time, τ_α , is then computed using Kramers' theory by incorporating both local and collective barrier contributions as^{7,11,12,24–39}

$$\frac{\tau_\alpha}{\tau_s} = 1 + \frac{2\pi}{\sqrt{K_0 K_B}} \frac{k_B T}{d^2} \exp\left(\frac{F_B + F_e}{k_B T}\right), \quad (5)$$

where τ_s is a short-timescale prefactor related to the combined effects of the binary collision rate and weak-caging effects, and

$K_B = \left| \partial^2 F_{\text{dyn}}(r)/\partial r^2 \right|_{r=r_B}$ is the absolute curvature at the barrier position of the dynamic free energy. The explicit form of τ_s was provided in previous studies.^{26–29}

Equation (5) yields τ_α as a function of volume fraction. However, direct comparison with the experimental data requires mapping density to temperature. Originally, Mirigian and Schweizer proposed a thermal mapping framework for rigid molecular liquids,^{29,31} which uses a more thermodynamic basis. This approach is based on reproducing the experimental long-wavelength dimensionless density fluctuation amplitude, $S_0^{\text{expt}}(T) = \rho k_B T \kappa_T$, where κ_T is the isothermal compressibility. This quantity captures nanoscale density fluctuations. This quantity is a sensitive probe of nanoscale density fluctuations and can be derived from the experimental equation of state (EOS). Within the hard-sphere model using the Percus–Yevick closure, S_0 is given analytically as $\frac{(1-\Phi)^4}{(1+2\Phi)^2}$. By equating this expression with an empirical fit to EOS data,^{24–34}

$$S_0^{\text{HS}} = \frac{(1-\Phi)^4}{(1+2\Phi)^2} \equiv S_0^{\text{expt}}(T) \approx \frac{1}{N_s} \left(-A + \frac{B}{T} \right)^{-2}, \quad (6)$$

where A and B are empirically derived EOS parameters representing entropy and cohesive energy contributions, respectively, and N_s is the number of interaction sites per molecule such as chemical groups of $-C-$, $-CH-$, $-O-$, $-NH-$, and so on. From this, one obtains the effective hard-sphere packing fraction as a function of temperature as^{31,32}

26 July 2025 10:38:18

$$\Phi_{\text{eff}}(T) = 1 + \sqrt{S_0^{\text{expt}}(T)} - \sqrt{S_0^{\text{expt}}(T) + 3\sqrt{S_0^{\text{expt}}(T)}}. \quad (7)$$

This approach requires knowledge of molecular-specific parameters (A , B , and N_s). Once $\Phi_{\text{eff}}(T)$ is obtained, we can calculate $g(r)$ and $S(q)$ and, hence, $F_{\text{dyn}}(r)$ and the temperature dependence of the structural relaxation time $\tau_a(T)$ without any adjustable parameters.

While the original ECNLE theory has successfully predicted glassy dynamics, across more than 14 decades, in the relaxation time for diverse molecular and polymeric glass-formers,^{24–34} it assumes a universal coupling between local cage dynamics and long-range elasticity. This assumption limits its ability to describe the significant variability in dynamic fragility observed among chemically distinct polymers. To overcome this limitation, Xie and Schweizer introduced a modified ECNLE approach incorporating a non-universal coupling between local cage-scale hopping and the longer-range collective elastic distortions.⁴⁶ Specifically, they assumed that the microscopic jump distance required for a local cage escape should be chemically specific and influenced by nanometer-scale conformational transitions and monomer-level interactions. This chemical specificity was implemented through a single material-dependent parameter a_c , which scales the collective elastic barrier in Eq. (4) as $F_{e,\text{new}} \equiv a_c F_e$. From this, Eq. (5) becomes

$$\frac{\tau_a}{\tau_s} = 1 + \frac{2\pi}{\sqrt{K_0 K_B}} \frac{k_B T}{d^2} \exp\left(\frac{F_B + a_c F_e}{k_B T}\right). \quad (8)$$

The extended theory significantly improves simultaneous predictions for the glass transition temperature, dynamic fragility, and the temperature dependence of τ_a .

Despite these improvements, practical application of the extended theory is limited by the challenge of accurately determining the EOS parameters and N_s for long-chain or chemically complex polymers. To address this limitation, we proposed an alternative thermal mapping in our prior works,^{35–37,39} based on the principle of thermal expansion, which is

$$T \approx T_g + \frac{\Phi_g - \Phi}{\beta \Phi_0}, \quad (9)$$

where $\beta \approx 12 \times 10^{-4} \text{ K}^{-1}$ is an effective thermal expansion coefficient (TEC), $\Phi_0 = 0.5$ is a characteristic packing fraction, and Φ_g corresponds to the volume fraction at the glass transition state defined when $\tau_a(\Phi_g) = \tau_a(T_g) = 100 \text{ s}$. Obviously, Φ_g depends on the magnitude of the material-specific parameter a_c , which scales the collective elastic barrier. In the case of $a_c = 1$, the glass transition packing fraction is set to $\Phi_g = 0.6157$, consistent with previous studies.^{35–37,39} Our simplified thermal mapping in Eq. (9) provides a simpler and more practical alternative to the original mapping approach.^{24–34,46} The original mapping approach is physically rigorous, and a recent work⁴⁷ has linked key thermodynamic parameters to the fragility and glass transition temperature of molecular and polymer liquids. However, this method requires detailed input such as A , B , and N_s , which are often difficult to determine experimentally, especially for complex or novel polymer systems.^{35–37,39,47} In contrast, our method is easier to apply to large

datasets or systems lacking full thermodynamic characterization. Despite its simplicity, this mapping still reflects the essential relationship between T_g and thermal expansion as reported in prior studies.^{48,49} This agreement reinforces the view that thermal expansivity encodes essential molecular-scale information governing the glass formation. The T_g value can be determined through DSC and/or BDS experiment, or alternatively estimated via computational simulations or machine learning,^{13–22,37,39,50} as indicated in Fig. 1.

B. Machine learning calculations

To implement the thermal mapping described in Eq. (9) to investigate a new polymer, machine learning models can be exploited to predict the glass transition temperature. In this work, we utilize a large-scale dataset compiled in Ref. 18, which contains 7174 polymers with experimentally measured T_g values and their corresponding chemical structures encoded using the Simplified Molecular Input Line Entry System (SMILES) notation. The dataset spans a wide range of T_g values from 134 to 768 K with an average of approximately 415.1 K. To uncover the structure–property relationships and predict the T_g values for new polymers, we apply four widely used ML algorithms including Gaussian Process Regression (GPR), Extra Trees Regression (ETR), Extreme Gradient Boosting (XGB), and Random Forest Regression (RFR). These models were selected for their proven ability to handle high-dimensional, non-linear relationships commonly observed in polymer datasets. Tree-based ensemble methods provide interpretability through feature importance measures,^{18,39} while GPR is well-suited for capturing complex trends with limited data through its probabilistic framework.^{13,51} For the GPR model, we follow the implementation procedure outlined in Ref. 51. However, for tree-based models, we adopted the methodology detailed in Ref. 39.

To evaluate model performance and quantify the predictive relationship between polymer structure and T_g , we trained the models using features derived from SMILES representations. The dataset was randomly split into training and testing subsets using an 80:20 ratio. The model accuracy was assessed using the coefficient of determination (R^2) and the root mean square error (RMSE) defined as

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (11)$$

where y_i and \hat{y}_i are the observed and predicted T_g values for the i th polymer, \bar{y} is the mean of the observed values, and n is the number of data points. Higher R^2 values and lower RMSE values indicate better model performance and predictive accuracy.

III. RESULTS AND DISCUSSION

Figure 2 compares experimental structural relaxation times, $\tau_a(T)$, with predictions from the ECNLE theory [Eqs. (1)–(5), and (9)]

26 July 2025 10:38:18

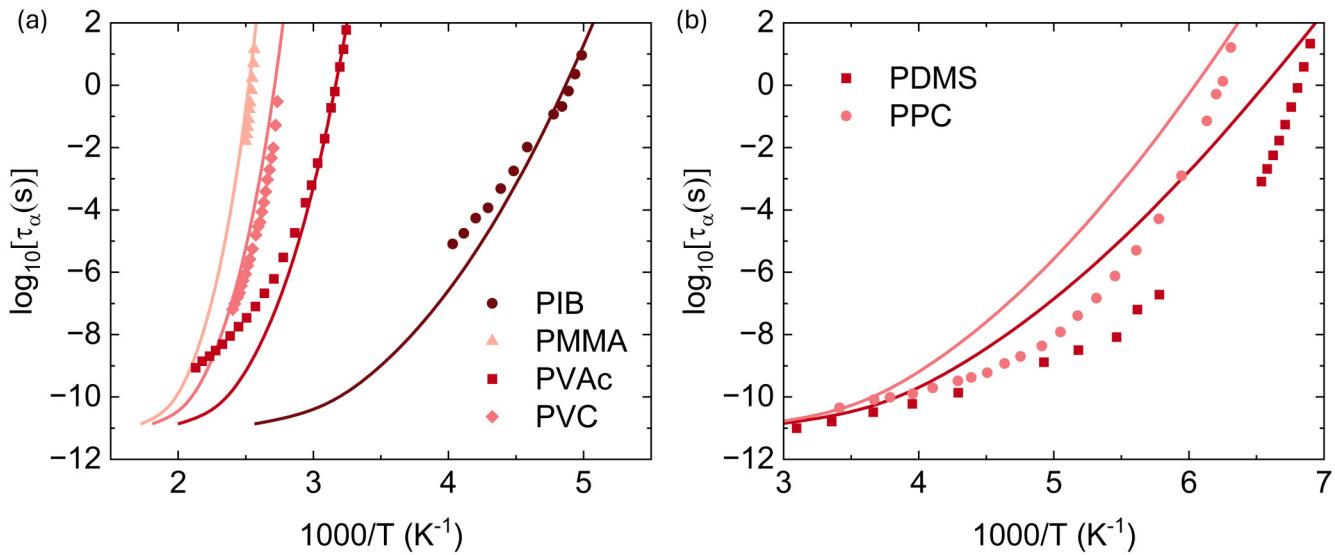


FIG. 2. Temperature dependence of the logarithm of the structural relaxation time of (a) polyisobutene (PIB), poly(methyl methacrylate) (PMMA), polyvinyl acetate (PVAc), and polyvinyl chloride (PVC), and (b) polydimethylsiloxane (PDMS) and polypropylene carbonate (PPC). Data points and solid curves correspond to the BDS experimental data in prior works^{54–58} and ECNLE calculations, respectively. The T_g values are determined using the BDS data.

for several representative polymers. As shown in Fig. 2(a), the theoretical calculations quantitatively reproduce the experimental data over a broad temperature range for PIB, PMMA, PVAc, and PVC. This is consistent with previous observations for many amorphous materials.^{35–37,39} This agreement validates the ability of the ECNLE framework to capture the essential features of glassy dynamics without any adjustable parameters. These polymers have also been previously studied using the $S_0(T)$ -based mapping with Kuhn segment-level coarse-graining in Refs. 31 and 46, where the ECNLE theory was able to quantitatively describe fragility and relaxation over a wide temperature range. Our simplified thermal expansion-based mapping provides a practical alternative in cases where detailed thermodynamic input is not available and still produces accurate results for these polymers. Our approach complements previous methods by trading some theoretical rigor for broader computational applicability.

In contrast, substantial deviations between the theoretical predictions and the experimental data are observed for PDMS and PPC as seen in Fig. 2(b). Although an accurate prediction of the glass transition temperature is essential to initiate thermal mapping, T_g alone does not fully dictate the temperature dependence of structural relaxation. These discrepancies likely arise from specific molecular characteristics including intramolecular flexibility, side-chain dynamics, and specific intermolecular interactions. To account for these effects, we adopt the approach proposed by Xie and Schweizer by introducing a material-specific parameter a_c that adjusts the importance of the collective elastic contribution to the total barrier ($F_B + a_c F_e$).^{35,46} Unlike thermodynamic approaches that rely on energy landscapes or packing structures, this modification directly captures how collective molecular motions slow down particle movement. The value of a_c is treated as an adjustable parameter to obtain the best fit to the experimental data.^{35,46}

Figure 3 shows the ECNLE calculations [Eqs. (1)–(4), (8), and (9)] for PDMS and PPC using different values of the a_c parameter and compares the results with experimental data. Although tuning a_c leads to better agreement, the predictions still fall short of quantitatively capturing the relaxation times across the full temperature range. Remarkably, the earlier ECNLE work assuming $a_c = 1$,³¹ when using the more detailed $S_0(T)$ -based thermal mapping and Kuhn segment coarse-graining, showed that despite underestimating the T_g value of PDMS, the model successfully captured key features such as the molecular weight dependence of T_g [normalized by its infinite molecular weight limit, $T_g(\infty)$] and the temperature dependence of τ_α in an Angell plot. In contrast, when our results for PDMS and PPC in Fig. 3 are replotted in the Angell representation (see Fig. S1 in the [supplementary material](#)), the theoretical curves do not quantitatively agree with the experimental data. This comparison suggests that the ECNLE calculations based on the simplified thermal expansion mapping [Eq. (9)] are not sufficient to fully capture the relaxation dynamics of polymers with complex molecular structures. While an accurate estimate of T_g is important, it does not guarantee a reliable prediction of the temperature-dependent behavior of structural relaxation. Further theoretical development is needed.

In recent studies,^{52,53} Schweizer and coworkers employed the Polymer Reference Interaction Site Model (PRISM) theory to calculate the static structure factor and direct correlation function for discrete semiflexible polymer chains. This approach explicitly accounts for chain connectivity and conformational stiffness, which are important features of polymer dynamics. The resulting structural information is then used to construct the dynamic free energy and compute relaxation times following the same procedures as in the ECNLE theory based on the hard-sphere fluid model discussed

26 July 2025 10:38:18

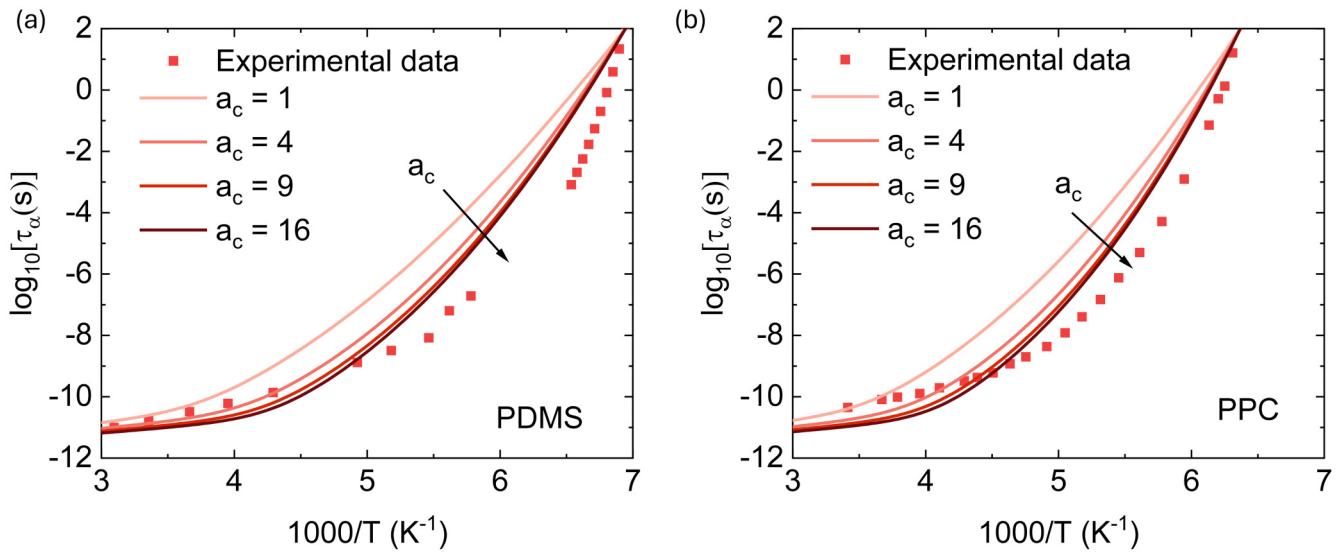


FIG. 3. Temperature dependence of the logarithm of the structural relaxation time of (a) PDMS and (b) PPC. Data points and solid curves correspond to the BDS experimental data in prior works^{54,59} and ECNLE calculations with various values of a_c , respectively. The T_g values are determined using the BDS data.

26 July 2025 10:38:18

in Sec. II. As shown in Fig. S1 of the [supplementary material](#), the predicted temperature dependence of the structural relaxation time from both the hard-sphere and polymer-specific models shows excellent agreement, even for varying chain lengths. This result suggests that the hard-sphere model, despite its simplicity, effectively captures the essential physics of glassy relaxation in polymer

melts and remains a robust and practical choice for providing an alternative method to the S_0 mapping idea, which both can be combined with the ENCLE theory, when detailed structural data are not available.

An important assumption in our simplified thermal mapping [Eq. (9)] is that the TEC of polymers remains nearly constant. This

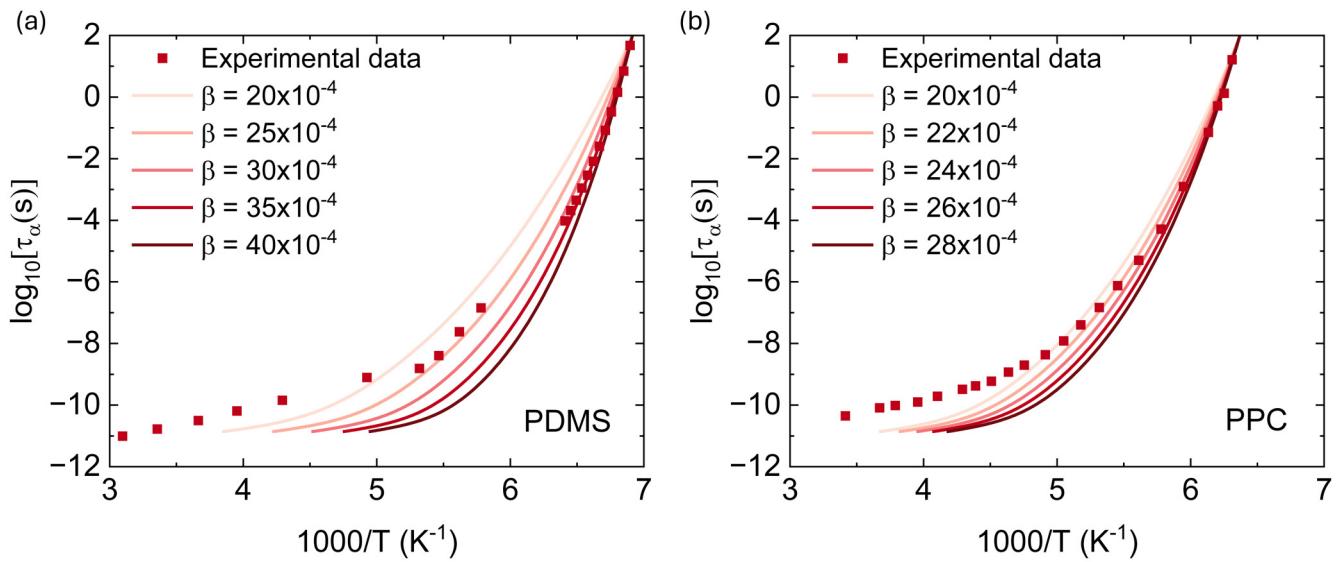


FIG. 4. Temperature dependence of the logarithm of the structural relaxation time of (a) PDMS and (b) PPC. Data points and solid curves correspond to the BDS experimental data in prior works^{54,59} and ECNLE calculations with various values of β in Eq. (9), respectively. The T_g values are determined using the BDS data.

assumption is well-supported by the experimental data reported in Ref. 49, where polymers with $T_g > 160$ K have remarkably similar thermal expansion behavior. Such consistency suggests that, above this T_g threshold, the volumetric response of polymers is governed by general thermodynamic constraints and is relatively insensitive to specific chemical structures. However, for polymers with lower glass transition temperatures ($T_g < 160$ K), including PDMS and PPC, the situation appears to deviate from this trend. These materials exhibit significantly greater chain flexibility and lower cohesive energy densities, which are expected to result in enhanced thermal expansivity. As a result, although the TEC in this low- T_g regime is still treated as constant to a zeroth-order approximation, it is hypothesized to be greater than that of high- T_g polymers. This hypothesis provides a plausible explanation for the discrepancies

observed in ECNLE predictions for PDMS and PPC, when using the standard thermal mapping parameters.

To test this hypothesis and assess its impact on ECNLE predictions, we perform new calculations for PDMS and PPC. Using Eq. (5) to compute $\tau_\alpha(\Phi)$, we re-apply the thermal mapping in Eq. (9) with an increased value of β to mimic the anticipated higher TEC in these systems. As shown in Fig. 4, this adjustment yields ECNLE predictions of $\tau_\alpha(T)$ that are in excellent quantitative agreement with the experimental BDS data. The predictive capability of the ECNLE theory is substantially enhanced relative to the case employing a universal β value. These results strongly support the notion that low- T_g polymers have significantly greater thermal expansion and validate the need for an adjusted mapping in such systems. However, due to the lack of direct experimental

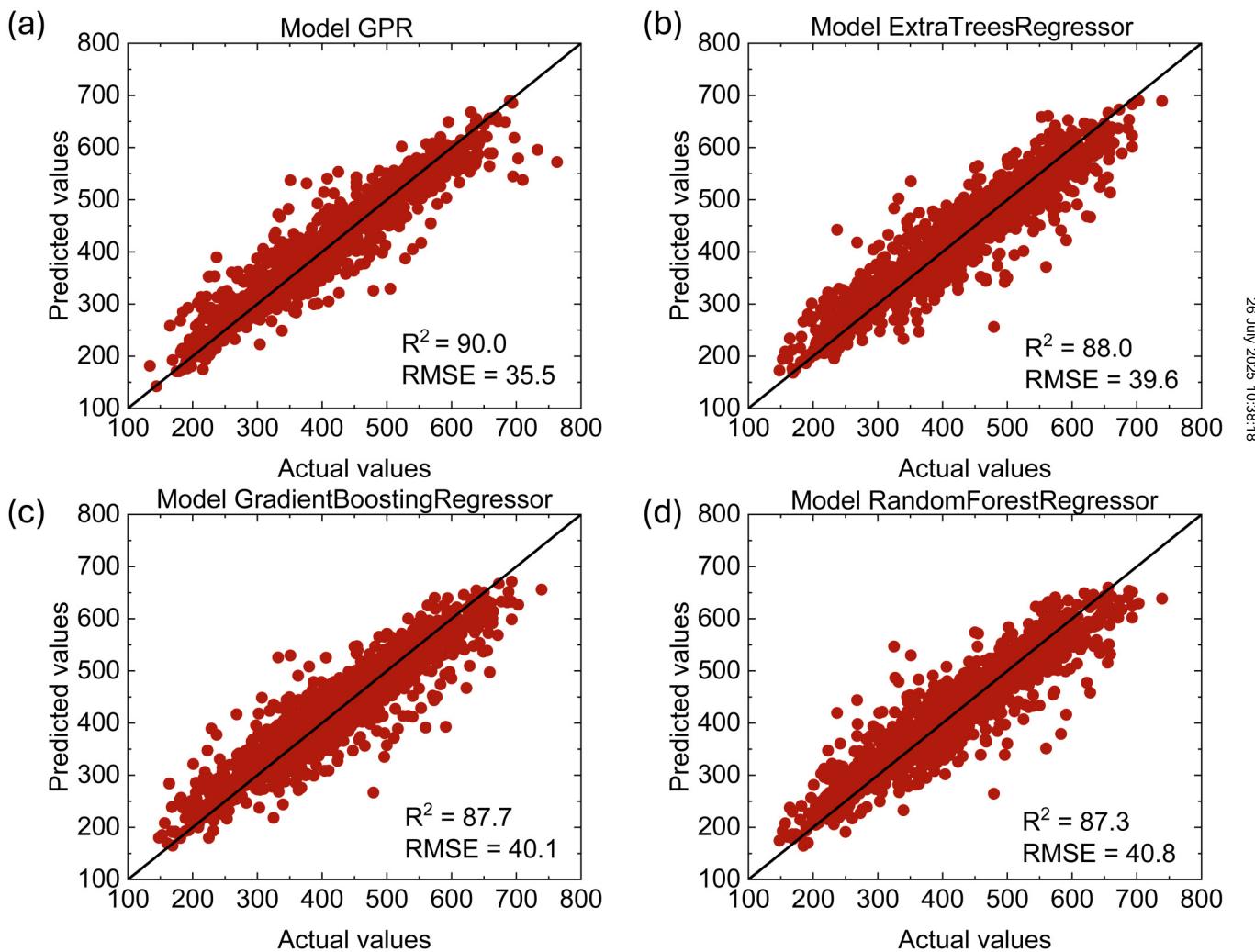


FIG. 5. Predictive performance of four regression models on the testing dataset for polymer T_g prediction: (a) Gaussian Process Regression, (b) Extra Trees Regression, (c) Extreme Gradient Boosting, and (d) Random Forest Regression.

measurements of TEC in this regime, this remains a working hypothesis. Future experimental investigations aimed at accurately quantifying the thermal expansion behavior of low- T_g polymers would be critical to confirm this assumption. To partially assess the plausibility of our chosen value, in Ref. 46, $A = 0.675$, $B = 1057$, and $N_s = 12.6$ for PDMS, we calculate $\Phi_{eff}(T)$ based on Eq. (6) and find that the slope of $\Phi_{eff}(T)$ is approximately $7.5 \times 10^{-4} \text{ K}^{-1}$ at high temperatures and around $10 \times 10^{-4} \text{ K}^{-1}$ at lower temperatures. It corresponds to the effective thermal expansion coefficient (β in our thermal mapping) in the range of $15\text{--}20 \times 10^{-4} \text{ K}^{-1}$, which is reasonably close to our employed value. Consequently, as a zeroth-order approximation, we adopt a thermal expansion coefficient of $\beta = 24 \times 10^{-4} \text{ K}^{-1}$, which is twice the universal value previously applied to all materials,^{7,11,12,24–39} for polymers with predicted T_g values below 160 K. This comparison provides further justification that our use of a higher β value for low- T_g polymers is not arbitrary, but rather grounded in the available experimental data.

The results presented above indicate that ECNLE calculations associated with the experimental T_g values yield excellent quantitative agreement with the BDS data. This raises an important question: can comparable predictive accuracy be achieved when substituting experimental T_g values with those obtained from the machine learning models? To address this, we trained four regression models (GPR, ETR, XGB, and RFR) on the polymer dataset reported in Ref. 18. As shown in Fig. 5, GPR achieved the highest predictive performance with an R^2 value of 90% and an RMSE of approximately 35.5 K. This outperforms the previously reported models applied to the same dataset (Table I), where the best R^2 was 86.9% and most RMSE values exceeded 40 K. Although ETR, XGB, and RFR in our study also surpassed their counterparts in previous works,¹⁸ their R^2 values remained below 89%. Increasing the training set generally leads to modest improvements in model accuracy, as shown in Table SI (supplemental material). Among all models tested, GPR consistently exhibits the best predictive performance in different train-test splits with $R^2 > 89\%$ and RMSE below 37 K.

To further assess the accuracy of our machine learning-assisted ECNLE predictions of glass transition temperature, we compare them with results from the previous ECNLE studies that employed the $S_0(T)$ -based thermal mapping as discussed in Refs. 31 and 46. In Ref. 31, Schweizer and coworkers employed the ECNLE theory in conjunction with the $S_0(T)$ -based thermal mapping [Eq. (7)] and assumed $a_c = 1$ to predict T_g values for 12 polymers. Their approach yielded RMSE of 46.32 K and R^2 of 76.34% against experimental T_g values. In a later study,⁴⁶ they introduced a specific chemistry parameter a_c to improve the predictive accuracy of both T_g and the dynamic fragility of 17 polymers and achieved RMSE ≈ 33.1 K and $R^2 \approx 86\%$. In comparison, our machine learning approach obtains similar or better predictive accuracy without the need for adjustable parameters and is trained on a much larger and more chemically diverse polymer dataset. Although the ECNLE theory provides important physical insights and captures the dependence of molecular weight on T_g , our data-driven approach provides an alternative to link the ECNLE theory to real materials, especially for high-throughput applications or systems that lack detailed thermodynamic input.

Note that certain specialized approaches in prior studies achieved even higher predictive accuracy but they have inherent

TABLE I. Summary of datasets, ML and DL models, and their corresponding R^2 and RMSE for the glass transition temperature prediction in previous works. The ML/DL models include Gaussian Process Regression (GPR), Graph Convolutional Network (GCN), Convolutional Neural Network (CNN), Generative Adversarial Networks (GANs), Molecular Structural Regularized Graph Convolutional Network with Reinforcement Learning (MSRGCN-RL), Extreme Gradient Boosting (XGB), Ridge Regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Regression (SVR), Decision Tree (DT), AdaBoost (AB), K-Nearest Neighbors (KNN), Random Forest Regression (RFR), Light Gradient Boosting Machine (LGB), Histogram-Based Gradient Boosting (HGB), Extra Tree Regression (ETR), Artificial Neural Network (ANN), Recurrent Neural Network (RNN), and Graph Convolutional Neural Network (GCNN).

Size of data	ML/DL model	R^2	RMSE	Reference
5076	GPR	...	35.3	13
300	GCN	88	19.4	14
>400	CNN	<80	49	15
	GANs	<92	31	
1200	MSRGCN-RL	92.76	12.07	16
800	XGB	97.10	0.736	17
	RR	84.49	2.542	
	LASSO	84.48	2.548	
	SVR	96.51	1.057	
7174	DT	73.4	57.299	18
	SVR	86.1	41.36	
	AB	69.0	61.920	
	KNN	82.9	45.963	
	XGB	86.2	41.249	
	RFR	82.5	46.507	
	LGB	85.9	41.773	
	HGB	86.1	41.371	
	ETR	86.9	40.286	
1437	KNN	71.0	...	19
	SVR	68.9	...	
	XGB	77.4	...	
	ANN	79.0	...	
	RNN	76.3	...	
60	GPR	...	2.2762	20
6923	CNN	83	48.45	21
	RNN	81	45.81	
	GCNN	88	38.82	
902	SVR	77	...	22
~7500	GCN	90	38.1	23

26 July 2025 10:38:18

limitations. Zhan and his coworkers¹⁷ attained an exceptional predictive accuracy ($R^2 \approx 97\%$) using the XGB algorithm combined with data augmentation techniques. However, their dataset was constructed exclusively from simulation data which lack the complexity and variability observed in experimental measurements. Consequently, their approach may not yield comparable accuracy when applied to experimental data. Similarly, Dong *et al.*¹⁶ integrated graph convolutional

networks and reinforcement learning to achieve a predictive accuracy of nearly 93%. However, this hybrid method is considerably more complex and their dataset primarily consisted of structurally similar polyimides. Clearly, the dataset lacks the chemical diversity present in our broader dataset. As a result, their high predictive accuracy might diminish if applied to a chemically diverse polymer dataset. These comparisons emphasize the strength of our GPR model in reliably capturing complex, nonlinear structure–property relationships across chemically diverse, experimentally derived polymer datasets due to its probabilistic formulation and flexibility in modeling smooth functions.

A recent study led by Glova and Karttunen⁵⁰ presents a machine learning framework that utilizes dimensionality reduction

of MD simulation data to estimate the glass transition temperature of polymers. Their method combines nonlinear diffusion maps, principal component analysis, and Gaussian mixture modeling to determine the T_g value from the radial distribution functions, mean square displacements, relative square displacements, and dihedral angles. While this approach provides valuable physical insight into glassy dynamics from atomistic simulations, its reliance on system-specific and high-resolution MD trajectories restricts the applicability for high-throughput screening and the evaluation of unsynthesized polymers. In contrast, our study introduces a data-driven framework that predicts T_g directly from the chemical structure of the polymer using experimental data and composition-

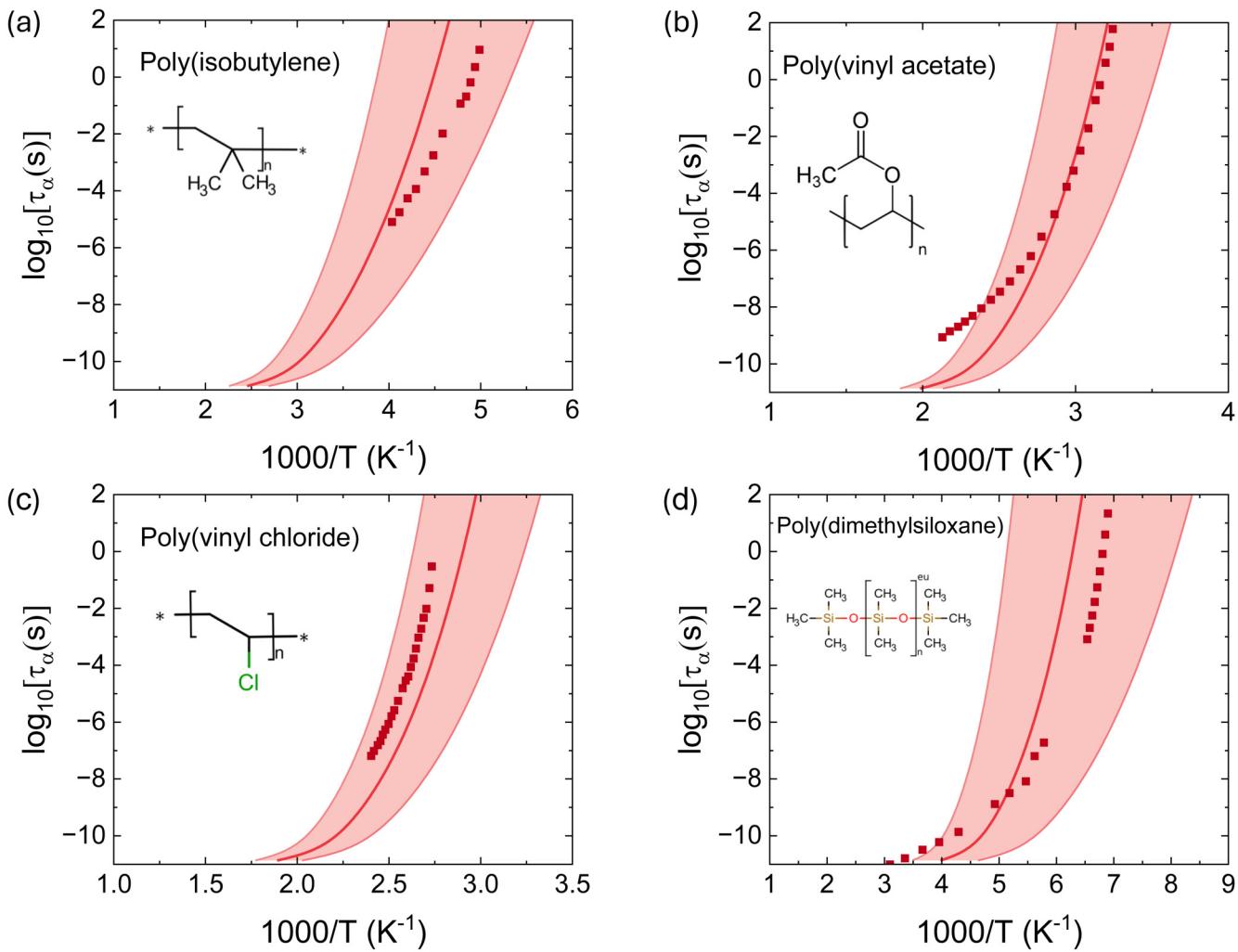


FIG. 6. Structural relaxation times plotted as a function of inverse temperature ($1000/T$) for four representative polymers: (a) poly(isobutylene) (PIB),⁵⁵ (b) poly(vinyl acetate) (PVAc),⁵⁷ (c) poly(vinyl chloride) (PVC),⁵⁸ and (d) poly(dimethylsiloxane).⁵⁴ Data points correspond to experimental data. The three solid curves correspond to ECNLE calculations incorporating the T_g values predicted by the GPR model. The central curve is the ECNLE prediction using the true GPR-predicted T_g , while the two adjacent curves correspond to the ECNLE results based on the T_g values adjusted by $\pm 10\%$ to account for the typical relative uncertainty associated with GPR predictions.

26 July 2025 10:38:18

based descriptors. Thus, this allows accurate and computationally efficient T_g prediction and is well-suited for material discovery and screening in broad chemical spaces.

Crucially, unlike earlier studies^{13–22,50} that focused exclusively on predictive accuracy, we incorporate machine-learning-derived T_g values into the ECNLE theory to compute the temperature dependence of the structural relaxation time. Figure 6 presents the theoretical and experimental structural relaxation times as a function of inverse temperature for four representative polymers. It is important to note that PIB, PVAc, PVC, and PDMS are not included in the training or the testing dataset used to develop the machine learning models. In each case, ECNLE predictions based on the T_g values obtained from Gaussian Process Regression are compared with the experimental BDS data. The central solid curve in each subfigure corresponds to the ECNLE calculations using the

true predicted GPR T_g , while the adjacent curves are the ECNLE predictions incorporating a relative uncertainty $\pm 10\%$ in T_g . Across all four polymers, the agreement between the experimental data and the ECNLE predictions is excellent. Remarkably, most experimental data points fall within the confidence interval (shaded band in Fig. 5). These findings clearly indicate that the GPR-predicted values T_g integrated into the ECNLE theory can quantitatively and accurately predict the temperature-dependent relaxation behavior in polymers. They further validate the effectiveness of combining machine learning with a physics-based theory to achieve reliable and predictive modeling. Our combined framework not only preserves predictive accuracy but also provides a powerful alternative to relying solely on experimentally measured T_g values, particularly for polymers that are computationally designed or experimentally inaccessible.

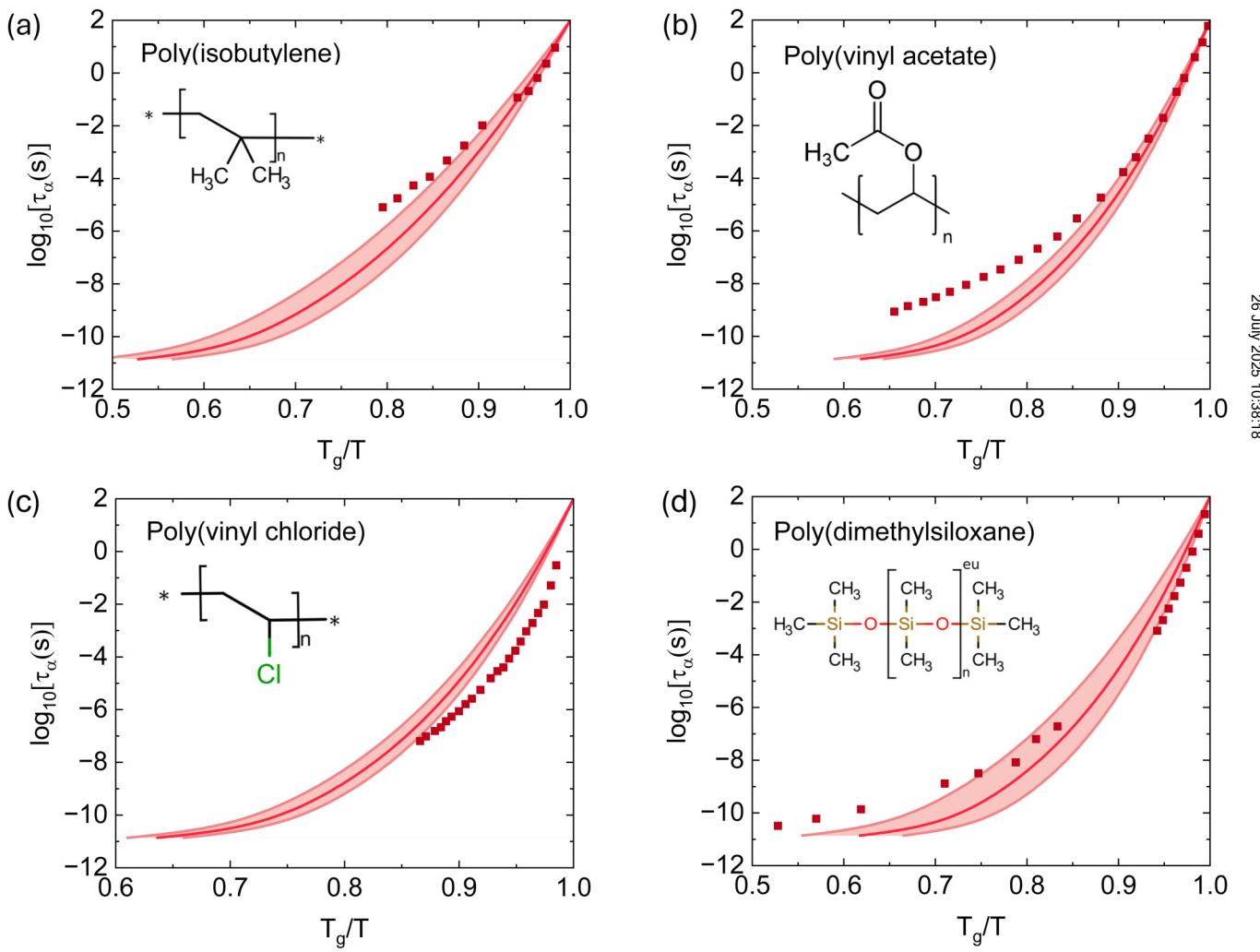


FIG. 7. Structural relaxation times plotted as a function of T_g/T for (a) PIB,⁵⁵ (b) PVAc,⁵⁷ (c) PVC,⁵⁸ and (d) PDMS.⁵⁴ Data points correspond to experimental data. Solid curves represent ECNLE calculations as shown in Fig. 6 using the GPR-predicted T_g values.

To further evaluate the accuracy of our theoretical framework, we replot the data of Fig. 6 in the Angell representation and show the results in Fig. 7. This representation facilitates a direct comparison of the temperature-dependent relaxation dynamics and dynamic fragility with the previously reported values in Refs. 31 and 46. From the calculations, the estimated dynamic fragilities for PIB, PVAc, PDMS, and PVC are approximately 61, 81, 80, and 88, respectively. These values are in reasonable agreement with the experimental data reported in Refs. 31 and 46, which are 46, 95, 79–100, and 191 for PIB, PVAc, PDMS, and PVC, respectively. While our calculations slightly underestimate the fragility in some cases, deviations are acceptable, especially considering that our model does not rely on any adjustable parameters. Notably, the ECNLE studies in Ref. 31 in conjunction with the $S_0(T)$ -based thermal mapping predicted fragilities of 85 for PIB, 95 for PVAc, 84 for PDMS, and 94 for PVC. Although this approach includes a more detailed thermodynamic input, its predictions still show noticeable deviations. The extended ECNLE theory in Ref. 46 further improves agreement by introducing a material-specific parameter a_c ; this enhancement comes at the cost of reduced generality. Since our model avoids such parameters, direct comparison should be interpreted with caution. For PVC, in particular, the fragility reported in Refs. 31 and 46 deviates significantly from the experimental value and reveals that even more sophisticated models can face limitations for certain systems. Overall, our simplified ECNLE framework remains effective in capturing the fragility trends across diverse polymers.

Based on the calculations and analysis presented above, we now establish a practical workflow for investigating the glassy dynamics of new or unsynthesized polymers. The process begins by providing the SMILES representation of a polymer as input to the trained Gaussian Process Regression model to predict the glass transition temperature. This predicted T_g is subsequently incorporated into the ECNLE theory to calculate the temperature dependence of the structural relaxation time. Specifically, the T_g value is used in the thermal mapping relation defined in Eq. (9). To account for the distinct thermal expansion behaviors observed in polymers with different T_g regimes, the effective thermal expansion coefficient is assigned based on the predicted T_g . For polymers with $T_g > 160$ K, we adopt $\beta = 12 \times 10^{-4} \text{ K}^{-1}$, while for those with $T_g < 160$ K, a larger value of $\beta = 24 \times 10^{-4} \text{ K}^{-1}$ is used to reflect the enhanced thermal expansivity expected in low- T_g systems. Although the approximation of a constant β simplifies the modeling approach and reduces the need for extensive experimental input, future work could examine how a temperature-dependent thermal expansion coefficient might improve the accuracy of ECNLE predictions. This approach provides a data-driven pipeline for predicting the glassy dynamics of polymers.

It is worth noting that a recent work⁶⁰ has shown that including thermodynamic and vibrational descriptors can enhance the ability of machine learning models to investigate the glassy dynamics in materials. Although our current approach uses only the chemical structure and composition-based features, it captures similar physical behavior through the use of the thermal expansion coefficient in our simplified thermal mapping. This parameter represents how the material responds to temperature changes and indirectly includes information related to molecular vibrations and

local softness. The quantitative agreement between our theoretical predictions and experimental data clearly supports the fact that thermodynamic and vibrational properties are important indicators of glassy relaxation.

IV. CONCLUSION

In summary, we have developed a unified framework that integrates machine learning with the ECNLE theory to enhance the predictive modeling of glassy dynamics in polymers. By training and validating several regression models on a large chemically diverse dataset of 7174 polymers, we have identified Gaussian Process Regression as the most accurate and generalizable model for predicting glass transition temperatures. Its performance is better than most previously reported approaches, in both accuracy and robustness, across a chemical space. Importantly, by integrating GPR-predicted T_g values into the ECNLE theory coupled with a simplified thermal mapping [Eq. (9)], we successfully captured the temperature dependence of structural relaxation times across various polymers. The theoretical predictions show excellent agreement with the experimental BDS data. For polymers with low predicted T_g values (below 160 K), we increase the thermal expansion coefficient β in the mapping equation from 12×10^{-4} to $24 \times 10^{-4} \text{ K}^{-1}$. This modification significantly improves the accuracy of the predicted relaxation behavior and suggests the potential need for future experimental studies to validate the thermal expansion properties in low- T_g materials.

SUPPLEMENTARY MATERIAL

The [supplementary material](#) include detailed benchmarking results for machine learning models trained to predict the glass transition temperature of polymers with performance metrics for different train-test splits. It also presents extended ECNLE calculations using chemically specific polymer models based on the PRISM theory and compares them with results from the hard-sphere fluid model to validate the simplified approach used in the main text.

ACKNOWLEDGMENTS

This research was funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant No. 103.01-2023.62.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Anh D. Phan: Conceptualization (lead); Funding acquisition (lead); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (lead); Supervision (lead); Validation (lead); Visualization (equal); Writing – original draft (lead); Writing – review & editing (lead). **Ngo T. Que:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal).

26 July 2025 10:38:18

Nguyen T. T. Duyen: Data curation (equal); Formal analysis (equal); Investigation (equal); Validation (equal). **Phan Thanh Viet:** Data curation (equal); Investigation (equal); Validation (equal); Visualization (equal). **Quach K. Quang:** Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal). **Baicheng Mei:** Data curation (equal); Investigation (equal); Methodology (equal); Resources (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- ¹C. A. Angell, *Science* **267**, 1924–1935 (1995).
- ²M. D. Ediger, *Annu. Rev. Phys. Chem.* **51**, 99–128 (2000).
- ³B. C. Hancock and G. Zografi, *J. Pharm. Sci.* **86**, 1–12 (1997).
- ⁴L. Berthier and G. Biroli, *Rev. Mod. Phys.* **83**, 587–645 (2011).
- ⁵H. Sillescu, *J. Non-Cryst. Solids* **243**(2–3), 81–108 (1999).
- ⁶M. D. Ediger and P. Harrowell, *J. Chem. Phys.* **137**, 080901 (2012).
- ⁷K. S. Schweizer and D. S. Simmons, *J. Chem. Phys.* **151**, 240901 (2019).
- ⁸R. Richert, *J. Phys.: Condens. Matter* **14**, R703 (2002).
- ⁹W. Götze, *Complex Dynamics of Glass-Forming Liquids: A Mode-Coupling Theory* (Oxford University Press, 2009).
- ¹⁰V. Lubchenko and P. G. Wolynes, *Annu. Rev. Phys. Chem.* **58**, 235–266 (2007).
- ¹¹A. Ghanekarade, A. D. Phan, K. S. Schweizer, and D. S. Simmons, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104398118 (2021).
- ¹²A. Ghanekarade, A. D. Phan, K. S. Schweizer, and D. S. Simmons, *Nat. Phys.* **19**, 800–806 (2023).
- ¹³H. D. Tran *et al.*, *J. Appl. Phys.* **128**, 171104 (2020).
- ¹⁴J. Hu, Z. Li, J. Lin, and L. Zhang, *ACS Appl. Mater. Interfaces* **15**, 54006 (2023).
- ¹⁵T. Nguyen and M. Bavarian, *Ind. Eng. Chem. Res.* **61**, 12690 (2022).
- ¹⁶C. Dong, D. Li, and J. Liu, *Langmuir* **40**, 18568 (2024).
- ¹⁷S. Zhan *et al.*, *Mater. Today Commun.* **40**, 110181 (2024).
- ¹⁸M. J. Uddin and J. Fan, *Polymers* **16**, 1049 (2024).
- ¹⁹J. F. Fatriansyah, B. D. P. Linuwih, Y. Andreatno, I. S. Sari, A. Federico, M. Anis, S. N. Surip, and M. Jaafar, *Polymers* **16**, 2464 (2024).
- ²⁰Y. Zhang and X. Xu *Heliyon* **6**, e05055 (2020).
- ²¹L. Tao, V. Varshney, and Y. Li, *J. Chem. Inf. Model.* **61**, 5395 (2021).
- ²²G. M. Casanola-Martin *et al.*, *Commun. Chem.* **7**, 226 (2024).
- ²³K. Hickey, J. Feinstein, G. Sivaraman, M. MacDonell, E. Yan, C. Matherson, S. Coia, J. Xu, and K. Picel, *Comput. Mater. Sci.* **238**, 112933 (2024).
- ²⁴A. D. Phan and K. S. Schweizer, *ACS Macro Lett.* **9**, 448–453 (2020).
- ²⁵A. D. Phan and K. S. Schweizer, *Macromolecules* **52**, 5192–5206 (2019).
- ²⁶A. D. Phan and K. S. Schweizer, *Macromolecules* **51**, 6063–6075 (2018).
- ²⁷A. D. Phan and K. S. Schweizer, *J. Chem. Phys.* **150**, 044508 (2019).
- ²⁸S. Mirigian and K. S. Schweizer, *J. Chem. Phys.* **140**, 194506 (2014).
- ²⁹S. Mirigian and K. S. Schweizer, *J. Chem. Phys.* **140**, 194507 (2014).
- ³⁰S. Mirigian and K. S. Schweizer, *J. Phys. Chem. Lett.* **4**, 3648–3653 (2013).
- ³¹S. Mirigian and K. S. Schweizer, *Macromolecules* **48**, 1901–1913 (2015).
- ³²A. D. Phan and K. S. Schweizer, *J. Phys. Chem. B* **122**, 8451–8461 (2018).
- ³³S. Mirigian and K. S. Schweizer, *J. Chem. Phys.* **141**, 161103 (2014).
- ³⁴S. Mirigian and K. S. Schweizer, *J. Chem. Phys.* **143**, 244705 (2015).
- ³⁵A. D. Phan, K. Wakabayashi, M. Paluch, and V. D. Lam, *RSC Adv.* **9**, 40214–40221 (2019).
- ³⁶A. D. Phan, A. Zaccione, V. D. Lam, and K. Wakabayashi, *Phys. Rev. Lett.* **126**, 025502 (2021).
- ³⁷A. D. Phan, D. T. Nga, N. T. Que, H. Peng, T. Norhourmour, and L. M. Tu, *Comput. Mater. Sci.* **251**, 113759 (2025).
- ³⁸A. D. Phan and K. S. Schweizer, *J. Chem. Phys.* **160**, 074902 (2024).
- ³⁹N. T. Que, A. D. Phan, T. Tran, P. T. Huy, M. X. Trang, and T. V. Luong, *Mater. Today Commun.* **45**, 112287 (2025).
- ⁴⁰B. Mei, T.-W. Lin, G. S. Sheridan, C. M. Evans, C. E. Sing, and K. S. Schweizer, *Macromolecules* **55**, 4159–4173 (2022).
- ⁴¹B. Mei, C. M. Evans, and K. S. Schweizer, *Macromolecules* **57**, 3242–3257 (2024).
- ⁴²T.-W. Lin, B. Mei, S. Dutta, K. S. Schweizer, and C. E. Sing, *Macromolecules* **58**, 1481–1497 (2025).
- ⁴³A. Ghosh, S. Samanta, S. Ge, A. P. Sokolov, and K. S. Schweizer, *Macromolecules* **55**, 2345–2357 (2022).
- ⁴⁴J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, London, 2006).
- ⁴⁵L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 3rd ed. (Permagon Press, London, 1975).
- ⁴⁶S.-J. Xie and K. S. Schweizer, *Macromolecules* **49**, 9655–9664 (2016).
- ⁴⁷B. Mei and K. S. Schweizer, *J. Phys. Chem. B* **128**, 11293–11312 (2024).
- ⁴⁸A. Zaccione and E. M. Terentjev, *Phys. Rev. Lett.* **110**, 178002 (2013).
- ⁴⁹P. Lunkenheimer, A. Loidl, B. Riechers, A. Zaccione, and K. Samwer, *Nat. Phys.* **19**, 694–699 (2023).
- ⁵⁰A. Glova and M. Karttunen, *J. Chem. Phys.* **161**, 184902 (2024).
- ⁵¹R.-R. Griffiths, J. L. Greenfield, A. R. Thawani, A. R. Jamasb, H. B. Moss, A. Bourached, P. Jones, W. McCorkindale, A. A. Aldrick, M. J. Fuchter, and A. A. Lee, *Chem. Sci.* **13**, 13541–13551 (2022).
- ⁵²Y. Zhou, B. Mei, and K. S. Schweizer, *J. Chem. Phys.* **156**, 114901 (2022).
- ⁵³B. Mei, T.-W. Lin, G. S. Sheridan, C. M. Evans, C. E. Sing, and K. S. Schweizer, *Macromolecules* **55**, 4159–4173 (2022).
- ⁵⁴B. Schmidtke, M. Hofmann, A. Lichtinger, and E. A. Rössler, *Macromolecules* **48**, 3005 (2015).
- ⁵⁵K. Kunal, M. Paluch, C. M. Roland, J. E. Puskas, Y. Chen, and A. P. Sokolov, *J. Polym. Sci., Part B: Polym. Phys.* **46**, 1390 (2008).
- ⁵⁶R. Casalini, L. Zhu, E. Baer, and C. M. Roland, *Polymer* **88**, 133 (2016).
- ⁵⁷M. Tyagi, A. Arbe, A. Alegria, J. Colmenero, and B. Frick, *Macromolecules* **40**, 4568–4577 (2007).
- ⁵⁸J. Colmenero, *Phys. A* **201**, 38–51 (1993).
- ⁵⁹P. Lunkenheimer and A. Loidl, *The Scaling of Relaxation Processes* (Springer, 2018), pp. 23–59.
- ⁶⁰Z.-Y. Yang, D. Wei, A. Zaccione, and Y.-J. Wang, *Phys. Rev. B* **104**, 064108 (2021).

26 July 2025 10:38:18

Supplementary Materials

Bridging Machine Learning and Glassy Dynamics Theory for Predictive Polymer Modeling

Anh D. Phan,^{1,2,*} Ngo T. Que,² Nguyen T. T. Duyen,¹ Phan Thanh Viet,³ Quach K. Quang,³ and Baicheng Mei⁴

¹*Faculty of Materials Science and Engineering, Phenikaa University, Hanoi 12116, Vietnam*

²*Phenikaa Institute for Advanced Study, Phenikaa University, Hanoi 12116, Vietnam*

³*Division of Physics, School of Education, Dong Thap University, Cao Lanh 870000, Vietnam*

⁴*School of Materials Science and Engineering, Beijing Institute of Technology, Beijing, 100081 China*

(Dated: July 7, 2025)

*Electronic address: anh.phanduc@phenikaa-uni.edu.vn

Table SI summarizes the test set performance of our regression models including Gaussian Process Regression (GPR), Extra Trees (ET), XGBoost (XGB), and Random Forest (RF) trained on a dataset of 7174 polymers. The calculations were performed with different train/test split ratios (60:40, 70:30, 80:20, and 90:10) to investigate the influence of training data size on predictive accuracy. The metrics used to evaluate the performance include the coefficient of determination (R^2) and root mean square error (RMSE in Kelvin). As the training set size increases, all models show slightly improved test performance. Notably, GPR consistently exhibits the highest accuracy across all data splits with R^2 values exceeding 89 % and RMSEs below 37 K even for the smallest training fraction. These results clearly indicate the stability and reliability of GPR in predicting T_g from polymer chemistry across varying data regimes.

Ratio training/testing	60/40	70/30	80/20	90/10
GPR	$R^2 = 89.5$ RMSE = 36.7 (K)	$R^2 = 90.1$ RMSE = 35.1 (K)	$R^2 = 90.0$ RMSE = 35.5 (K)	$R^2 = 91.2$ RMSE = 34.04 (K)
ET	$R^2 = 86.1$ RMSE = 42.4 (K)	$R^2 = 86.7$ RMSE = 40.6 (K)	$R^2 = 88.0$ RMSE = 39.6 (K)	$R^2 = 88.5$ RMSE = 38.8 (K)
XGB	$R^2 = 85.0$ RMSE = 43.9 (K)	$R^2 = 85.7$ RMSE = 29.7 (K)	$R^2 = 87.7$ RMSE = 40.1 (K)	$R^2 = 87.1$ RMSE = 28.0 (K)
RF	$R^2 = 85.8$ RMSE = 42.8 (K)	$R^2 = 86.7$ RMSE = 40.7 (K)	$R^2 = 87.3$ RMSE = 40.8 (K)	$R^2 = 88.7$ RMSE = 38.4 (K)

Table S I: Performance of machine learning models for predicting the glass transition temperature of polymers with different train/test split ratios.

Recent developments in the ECNLE theory [1, 2] have incorporated chemically specific polymer models to better capture the influence of chain connectivity and conformational stiffness on glassy dynamics. In this formulation, the static structure factor, $S(q)$, and direct correlation function, $C(q)$, are computed using the Polymer Reference Interaction Site Model (PRISM) for discrete semiflexible chains (or Koyama model) with the density ρ . Each polymer chain consists of N segments with diameter d , and a fixed bond length $l = 4d/3$ to represent of both chain flexibility and packing structure. These structural quantities are used to construct the dynamic free energy. From these, the local barrier and collective elastic barrier are computed using the same methodology as the ECNLE theory based on the hard-sphere fluid model. To enable quantitative comparison with experimental relaxation data, we apply a modified version of our simplified thermal mapping [Eq. (9) in the main text] to relate the density to temperature, which given by

$$T \approx T_g + \frac{\rho_g - \rho}{\beta \rho_0}, \quad (1)$$

where ρ_g is the density at the glass transition defined as $\tau_\alpha(\rho_g) = 100$ s, T_g is the glass transition temperature, $\beta = 12 \times 10^{-4} K^{-1}$ is still the effective thermal expansion coefficient, and $\rho_0 = 6\Phi_0/\pi \approx 0.955$ corresponds to the reference packing fraction. The value of ρ_g depends on the length of the chain.

We performed ECNLE calculations for PDMS polymer chains using the same setup and methodology as described in Refs. [1, 2] with $N = 4, 16, 64$, and 128 , and compare these results with predictions based on the hard-sphere fluid model. As shown in Fig. S1, the theoretical temperature dependence of structural relaxation time from both approaches exhibit excellent agreement across the entire temperature range. This strong consistency suggests that the hard-sphere-based ECNLE model remains a sufficiently accurate and broadly applicable framework for describing structural relaxation in polymer melts, even in the absence of detailed chain-specific modeling.

-
- [1] Y. Zhou, B. Mei, K. S. Schweizer, J. Chem. Phys. 156, 114901 (2022).
[2] B. Mei, T.-W. Lin, G. S. Sheridan, C. M. Evans, C. E. Sing, and K. S. Schweizer, Chem. Sci. 13, 13541-13551 (2022).

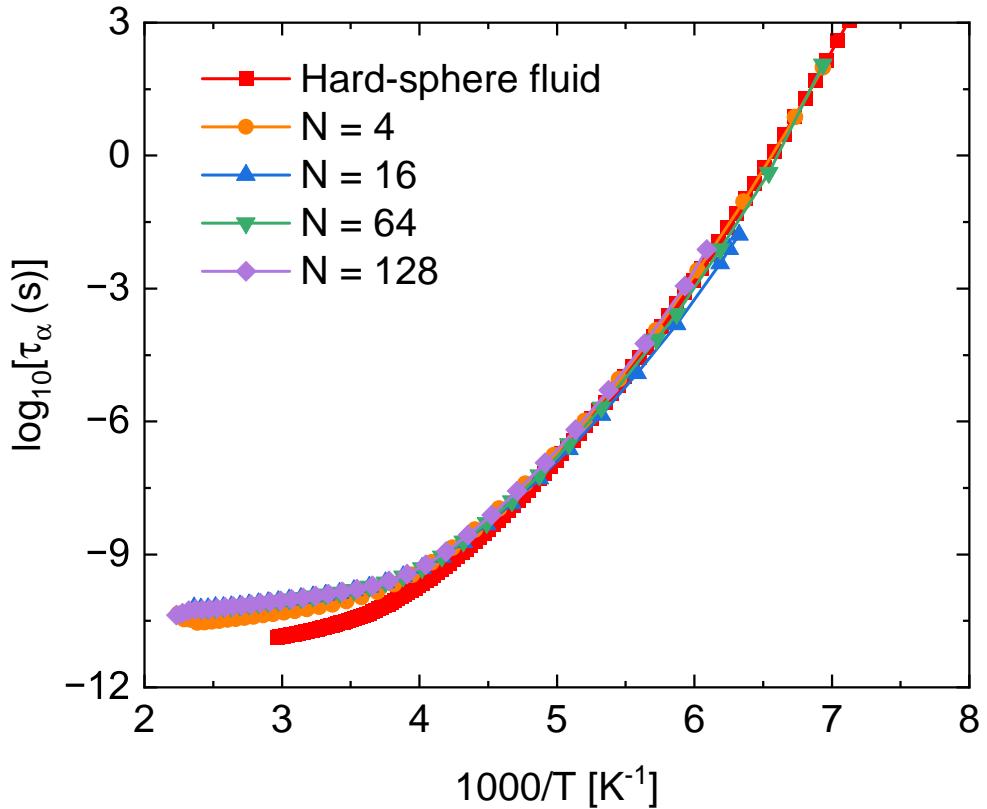


Figure S 1: Temperature dependence of the logarithm of the structural relaxation time calculated using ECNLE theory for both the hard-sphere fluid model and the chemically specific polymer model developed in Refs. [1, 2] with varying chain lengths ($N=4,16,64,128$).