Probing the Rheological Properties of Liquids Under Conditions of Elastohydrodynamic Lubrication Using Simulations and Machine Learning

J. C. S. Kadupitiya¹ · Vikram Jadhao¹

Received: 1 February 2021 / Accepted: 5 May 2021 / Published online: 27 May 2021
© The Author(s) 2021

Abstract

In elastohydrodynamic lubrication (EHL), the lubricant experiences pressures in excess of 500 MPa and strain rates larger than 10^5 s^{-1} . The high pressures lead to a dramatic rise in the Newtonian viscosity and the high rates lead to large shear stresses and pronounced shear thinning. The extraction of accurate rheological properties using non-equilibrium molecular dynamics simulations (NEMD) has played a key role in improving our understanding of lubricant flow in EHL conditions. However, the high dimensionality of the output data generated by NEMD simulations often makes a deeper interrogation of the link between molecular-scale features and rheological properties challenging. In this paper, we explore the use of machine learning to analyze and visualize the high-dimensional output data generated in typical NEMD simulations. We show that dimension reduction of NEMD simulation data describing the shear flow of squalane enables a clear visualization of the transition from Newtonian to non-Newtonian shear thinning with increasing shear rate and provides a reliable assessment of the correlation between shear thinning and the evolution in molecular alignment. The end-to-end atom pairs dominate the largest variations in pair orientation tensor components for low-pressure systems (0.1, 100 MPa) and provide the clearest separation of the orientation tensors with rate. On the other hand, the side atom pairs dominate the largest variation in the tensor components for the high-pressure systems ($P \geq 400 \text{ MPa}$) which exhibit an overall limited evolution in orientation tensors as a function of rate. Dimension reduction using all the six components of the orientation tensors of all 435 pairs associated with a squalane molecule shows that the decrease in viscosity with rate for low pressures is strongly correlated with changes in molecular alignment. However, for high pressures, shear thinning occurs at saturated orientational order.

Keywords Elastohydrodynamic lubrication · Rheology · Molecular dynamics simulations · Machine learning

1 Introduction

Many liquids when sheared exhibit a decrease in viscosity with the rise in shear strain rate [1, 2]. This shear thinning behavior depends on many factors such as thermodynamic conditions, strain rates, and the molecular structure of the liquid. Shear thinning and the scaling of the Newtonian viscosity with temperature and pressure are critical to the performance of liquids in many practical applications. For example, the Newtonian viscosity and shear thinning of fracturing fluids determine whether the fluid viscosity is

sufficiently large to reduce particle settling rates and sustain fractures of desired geometry against closure stresses imposed by the reservoir [3, 4]. In elastohydrodynamic lubrication (EHL), high sliding velocities of solid machine components produce strain rates $> 10^5 \text{ s}^{-1}$ and substantial shear thinning of lubricant films, which lowers the frictional stress [5–7]. Further, the high normal loads compress the lubricant in EHL contacts to pressures in excess of 500 MPa, which increases the Newtonian viscosity of the lubricant by orders of magnitude, preventing squeeze out and limiting wear.

Accurate models describing the rheological properties of liquids under EHL conditions are critical to predicting friction. Coupled with a fundamental molecular-level understanding of shear thinning, these models can also help guide the design of synthetic lubricants. However, developing these rheological models have proved to be challenging [5,

✉ Vikram Jadhao
vjadhao@iu.edu

¹ Intelligent Systems Engineering, Indiana University, Bloomington, IN 47408, USA

8, 9]. A key reason is the difficulty to reach the relevant pressures and strain rates in laboratory experiments. Specially designed rheometers have been able to exceed 1 GPa at 10^4 s^{-1} and reach shear stresses of up to $\sim 30 \text{ MPa}$ [5, 10–12], but may be affected by unmeasured temperature increases during shear at high rates [5]. Experiments on idealized EHL contacts can reach the higher rates, pressures, and shear stresses ($\sim 150 \text{ MPa}$) important in real devices [5, 7], but measure an average stress over contact regions with a range of local strain rates, pressures and temperatures [6, 13]. EHL conditions also present fundamental theoretical challenges as fluids are in a regime where they exhibit strongly non-linear viscoelastic behavior and poorly understood glassy rheology. While there is a general agreement on the qualitative changes in shear stress with increasing rate, there is an ongoing debate regarding the underlying mechanisms of shear thinning and the most suitable functional form of the rheological model that best describes liquid flows in EHL conditions [5, 6, 13].

Two phenomenological models at the center of this active debate are based on the work by Eyring [14, 15] and Carreau [16], respectively. The Eyring model assumes that flow occurs by thermal activation over a single characteristic energy barrier and predicts that shear stress σ rises as $\log(\dot{\gamma})$ at high strain rate $\dot{\gamma}$. In the Carreau model, σ increases as a power law, $\sigma \propto \dot{\gamma}^n$, with n typically around 0.5. This power-law behavior can arise from several different physical pictures of flow, including thermal activation over a broad range of energy barrier heights or a continuous increase in some type of order (e.g., molecular alignment) with increasing rate. Both of these competing models can be fit to the limited range of existing rheometer data on simple fluids that extends up to rates of about 10^4 s^{-1} . The two models predict dramatically different shear stresses and viscosities when extrapolated to rates that are important in EHL ($\dot{\gamma} > 10^5 \text{ s}^{-1}$), but inaccessible to rheometers [5]. Resolving the debate about flow behavior at these rates is of great practical and fundamental interest. In practical terms, it will yield accurate quantitative information about the macroscopic flow properties that are essential in predicting EHL friction and guiding synthetic lubricant design. From a fundamental perspective, it will advance our understanding of the molecular-level mechanisms of shear thinning and the nature of flow in glassy systems.

Non-equilibrium molecular dynamics (NEMD) simulations [17, 18] can probe shear flow under high strain rates ($\dot{\gamma} > 10^5 \text{ s}^{-1}$) which are challenging to access experimentally. These simulations have been used to investigate the shear thinning behavior of small-molecular liquids [19–22] as well as long-chain polymer molecules [23–26]. However, they are limited in probing the transport properties at lower rates due to increased computational costs associated with longer simulation times. Most of the early work on using

NEMD to study model lubricants considered systems at state points far from conditions prevalent in EHL and was limited to short simulation times and high strain rates accessible to the computers at the time [20, 27–30].

Increased computational power and the development of accurate all-atom force fields at high pressure has enabled recent flow studies under pressure and temperature conditions prevalent in EHL [7–9, 31–34]. Much of this work has been pioneered by Mark Robbins. Jadhao and Robbins [9, 31, 35] studied the nature of shear thinning in squalane by evaluating the shear stress, viscosity, and molecular order for a wide range of strain rates (10^5 – 10^{10} s^{-1}) and thermodynamic conditions (pressures: 0.1–1200 MPa; temperatures: 150–373 K) using NEMD simulations of a united-atom model of squalane [9, 31]. Their findings were consistent with available experimental data and, by extending the range of Newtonian viscosity η_N and rates to as low as 10^5 s^{-1} , revealed two distinct regimes of rheological behavior. At high temperatures and low pressures, where η_N of squalane is low, shear thinning viscosity was well fit by Carreau model, but with a scaling exponent n changing with η_N . As η_N increased above $\sim 1 \text{ Pa-s}$, squalane began to exhibit glassy flow and shear thinning was better described by the Eyring model over six orders of magnitude relative to η_N . For the high viscosity regime of experiments ($\eta_N > 10^3 \text{ Pa-s}$), the simulations and experiments were consistent with the Eyring equation over more than 8 orders of magnitude in strain rate. In this regime, using time-temperature superposition, simulation and experimental data were collapsed on the analytic solution of the Eyring model over almost 30 orders of magnitude in rate.

In a subsequent paper, Galvani Cunha and Robbins [32] employed a similar NEMD approach to study the shear response of 2,2,4-trimethylhexane using an all-atom force field developed by O'Connor, Andzelm, and Robbins for simulating shear flow at very high pressures [33]. They also observed a crossover from Carreau at low η_N to Eyring response at high η_N . More recently, Lin and Kedzierski employed the Eyring model to describe the NEMD simulation data for the rate-dependent viscosity of PEC6 at different temperatures and pressures over which η_N was changed by 4 orders of magnitude [34]. They found that the reduced viscosity vs rate data collapsed onto the analytic form of the Eyring equation. No universal Carreau equation could describe all the viscosity data for PEC6 as the power-law exponent n varied appreciably with temperature and pressure.

Shear thinning is often attributed to molecular ordering such as the alignment of chain-like molecules along the flow direction. By computing the ensemble-averaged components of the orientation order tensor associated with the pairs of end atoms, Jadhao and Robbins showed that while squalane molecules tend to align along the shear

flow direction, the degree of order saturated before the viscosity dropped by more than half a decade. Shear thinning continued to occur after the squalane chains have aligned and the decrease in viscosity extended to over six decades and over more than eight decades in rate. This regime was characterized with nearly constant molecular alignment and the associated flow was best described by the Eyring model.

The reliable extraction of accurate macroscopic rheological properties using NEMD simulations has played a key role in improving our understanding of lubricant flow in EHL conditions. Further, the computation of molecular-scale quantities such as average end-to-end molecular distances, alignment angles, and radii of gyration using the simulation data have provided key insights into the microscopic origins of changes in the rheological properties. However, the high dimensionality of the output data generated by the NEMD simulations often makes a deeper interrogation of the link between molecular-scale features and rheological properties challenging. For example, in the case of squalane, a more informative link between rheological properties and changes in the intramolecular structure can be established by examining pair orientation tensors associated with all the 435 distinct pairs of united atoms, with each tensor characterized by six non-trivial pair distance components. Visualizing this 435×6 high-dimensional information is difficult using conventional analysis tools. As NEMD simulations explore EHL fluids containing molecules of more complex shapes or mixtures of molecules of different chemical structures and lengths, new tools will be necessary to expedite the analysis of the high-dimensional simulation outputs aimed at identifying key molecular features (e.g., subset of atom pairs) that best classify evolution in molecular structure (e.g., alignment along the flow direction).

Machine learning (ML) techniques have the potential to address these needs and facilitate the analysis of information hidden in the high-dimensional data generated by standard NEMD simulations. Recent years have seen a surge in ML approaches to assist, enhance, or bypass MD simulations aimed at understanding material phenomena [36–38]. ML has been employed to rapidly scan assembly landscapes employing particle trajectories [36, 39], design efficient force fields [40], design integrators [41], and develop surrogate models to predict specific simulation outcomes bypassing part or all of the simulation [42–48]. Studies have demonstrated that ML methods can classify conformations of polymer chains and characterize structural relaxations in glassy materials [37, 49, 50]. However, with the exception of a few recent developments [51], there is a lack of utilization of ML techniques in examining the rheological behavior of liquids. The application of ML methods in the specific context of NEMD simulations related to tribological applications has not been explored to the best of our knowledge.

Fig. 1 Typical conformation of a squalane molecule at 293 K modeled using the united-atom model with carbons and hydrogens combined into united atoms (circles) that are connected by covalent bonds (lines). Red, green, and blue circles indicate carbon atoms with three, two, and one hydrogen atoms, respectively. Atoms are labeled with indices to facilitate the identification of pairs of atoms (Color figure online)

In this paper, we explore the use of ML tools such as dimension reduction methods to analyze and visualize the data generated in typical NEMD simulations. Using the extensive amount of simulation data generated for squalane, we show that ML methods can aid in the investigation and examination of the molecular-level mechanisms underlying the changes in the rheological properties. We show that the dimension reduction methods can enable a clear visualization of the transition from Newtonian flow to non-Newtonian shear thinning with increasing shear rate, and provide a reliable assessment of the correlation between shear thinning and the evolution in molecular alignment. Our findings suggest that the integration of ML methods with simulation and experimental approaches can further improve our understanding of lubricant flow behavior under EHL conditions.

The following section provides a brief overview of flow studies performed by Jadhao and Robbins [9]. Section 3 presents details of the simulation data preparation and dimension reduction methods utilized. Dimension reduction results for shear thinning in ambient conditions and under different higher pressures are given in Sect. 4. The final section presents discussion and conclusions.

2 Non-equilibrium Molecular Dynamics Simulations of Squalane

Squalane or 2,6,10,15,19,23-hexamethyltetracosane ($C_{30}H_{62}$) is a branched alkane with a C_{24} backbone and six methyl side groups. As shown in Fig. 1, the side groups are placed symmetrically about the backbone in a way that frustrates crystallization. Squalane has played an important role in the debate about rheological models of shear thinning under pressure and temperature conditions prevalent in EHL [5, 6, 8, 9, 13]. Due to its relatively simple chemical structure and small molecular size (end-to-end

distance ~ 2 nm), squalane has also been chosen for several computational studies of rheological properties of liquids [9, 20, 28, 30, 31]. Below we briefly summarize the methods and results of the flow studies performed by Jadhao and Robbins using NEMD simulations of squalane at 293 K [9, 31]; a detailed account can be found in previous publications [9, 31, 35].

Squalane is modeled using a united-atom (UA) potential based on the model developed by Mondello and Grest (MG) for branched alkanes [28, 52, 53]. In the UA description, hydrogens are lumped together with carbons into UA groups such as CH_3 , CH_2 , and CH . Each squalane molecule is represented by 30 UAs (Fig. 1). Within each molecule, the UAs interact via a harmonic bond-bending term, a harmonic bond-stretching term, a torsional potential, and a harmonic bending term to prevent umbrella inversion at tertiary carbon branch points. There is an additional non-bonded Lennard-Jones potential between UAs on different molecules or separated by at least four bonds on the same molecule.

In all cases, simulations are performed with LAMMPS [54] and used $N_m = 125$ squalane molecules. Periodic boundary conditions are used to prevent edge effects, and the temperature is maintained with a Nosé-Hoover thermostat. An initial state is created at $T = 293$ K and ambient pressure $P = 0.1$ MPa where diffusion is rapid and equilibrium is reached quickly. High-pressure states are generated by gradually compressing this state at constant T until P reached the desired value. Typical equilibration times are ≈ 100 nanoseconds and the simulation time step is 1 femtosecond.

Shear is imposed using the SLLOD equations of motion to study steady-state flow behavior [18]. The periodic box is sheared along the x -axis at a constant rate to impose the desired average strain rate $\dot{\gamma} = \delta v_x / \delta y$, with the velocity v_x along the x -axis and velocity gradient along the y -axis. For each $\dot{\gamma}$, the system is sheared until it reached steady state. The total strain required to reach steady state is of order 5 to 10. After reaching steady state, simulations are run for strains large enough to achieve the desired statistical accuracy in average quantities such as the stress and viscosity.

The rheological data from experiments and simulations are compared to fits from Eyring and Carreau models. The Eyring model is based on the idea that shear flow is a stress-biased thermally activated process [14, 15], that is, shear can only occur through molecular rearrangements that require activation over potential energy barriers with a single characteristic height. The model predicts that the shear stress σ rises sublinearly with strain rate according to the Eyring equation $\sigma = \sigma_E \sinh^{-1}(\eta_N \dot{\gamma} / \sigma_E)$ where $\sinh^{-1}$ is the inverse hyperbolic sine function, and $\sigma_E = k_B T / V^*$ is a parameter known as the Eyring stress that is related to the activation volume V^* . The Eyring equation predicts Newtonian behavior with a constant viscosity for $\sigma \ll \sigma_E$ and $\dot{\gamma} \ll \dot{\gamma}_E$, where

$\dot{\gamma}_E = \sigma_E / \eta_N$ is called the Eyring rate. The shear thinning behavior is given by the rate-dependent viscosity

$$\eta = \frac{\eta_N \dot{\gamma}_E}{\dot{\gamma}} \sinh^{-1}(\dot{\gamma} / \dot{\gamma}_E). \quad (1)$$

Limiting behavior for high rates sets in for $\dot{\gamma} \gg \dot{\gamma}_E$, where the stress rises logarithmically with rate, $\sigma \sim \log(\dot{\gamma})$, and the corresponding viscosity decreases as $\eta \sim \dot{\gamma}^{-1} \log(\dot{\gamma})$.

The Carreau model [16] belongs to the set of power-law fluid models which describe shear flow by assuming that there is a very broad distribution of barrier heights. These models are well suited to describe low viscosity fluids where molecules are in almost continual motion rather than hopping between cages formed by their neighbors. A power-law shear thinning behavior can also result where a continuous increase in some type of order with increasing rate leads to reductions in dissipation during flow. For example, individual molecules may stretch or align and groups of molecules may order into lines to reduce friction. The rate-dependent viscosity is given by the Carreau equation

$$\eta = \eta_N \left[1 + (\dot{\gamma} / \dot{\gamma}_0)^2 \right]^{\frac{n-1}{2}} \quad (2)$$

which has a specific form of the crossover from Newtonian behavior at $\dot{\gamma}$ below a characteristic rate $\dot{\gamma}_0$ to power-law behavior at $\dot{\gamma} \gg \dot{\gamma}_0$. In the high rate limit, the stress rises with rate as a power law, $\sigma \sim \dot{\gamma}^n$, where the exponent n is between 0 and 1, and the corresponding viscosity shows power-law shear thinning: $\eta \sim \dot{\gamma}^{-(1-n)}$.

We now summarize the results for shear stress, viscosity, and molecular alignment obtained by Jadhao and Robbins [9, 31]. The simulation data shown in Figs. 2 and 3 are the same as in their paper [9]. Figure 2a shows the plot of shear stress σ vs logarithm of rate $\dot{\gamma}$ at $T = 293$ K for pressures between 0.1 and 1200 MPa, focusing on values where there is experimental data [10, 11, 30]. Simulation results are shown along with fits to the Eyring and Carreau equations. For $P \leq 500$ MPa, simulations reached the Newtonian regime and the limiting η_N is used in the fits. For P between 636 and 955 MPa, fits are constrained to have the experimental value of η_N . No data are available to constrain η_N at 1200 MPa, so η_N is included in the fit.

High-pressure results ($P \geq 300$ MPa) from experiments and simulations are found to be consistent with the Eyring equation over up to 8 orders of magnitude in rate. At high rates, σ rises logarithmically with $\dot{\gamma}$ and the slope gives the Eyring stress σ_E , which is ≈ 9.3 MPa for $500 \leq P \leq 955$ MPa, and ≈ 9.6 MPa for 1200 MPa. Fits to simulations extending to rates as low as 10^5 s^{-1} are consistent with experiments at the same pressure that reach rates up to 10^4 s^{-1} and extend down into the Newtonian regime. As the pressure decreases, the range of rates where logarithmic behavior

Fig. 2 Pressure dependence of shear stress σ and viscosity η vs strain rate $\dot{\gamma}$ for squalane at $T = 293$ K. **a** Linear–log plot of σ vs $\dot{\gamma}$ at pressures P corresponding to the symbols defined in the legend. Symbols for $\dot{\gamma} \geq 10^5$ s $^{-1}$ correspond to simulation results obtained in [9], and symbols at lower rates are experimental values [10, 30]. Black dotted lines are Eyring model fits to simulation data and red dashed lines are Carreau fits. Fits are constrained to the limiting Newtonian viscosity from simulations for $P \leq 500$ MPa and from non-equilibrium experiments for P between 636 and 955 MPa. The Newtonian viscosity is fit at 1200 MPa. **b** η vs $\dot{\gamma}$ calculated from the stresses shown in Fig. 2a at P indicated in the legend of (a). Black dotted lines are Eyring fits and red dashed lines are Carreau fits (Color figure online)

is observed decreases. For $P < 200$ MPa ($\eta_N \lesssim 1$ Pa·s), the plots in Fig. 2a show clear curvature and are not well fit by the Eyring equation. In contrast, this low-pressure data are reasonably fit by the Carreau equation. The small deviations in the Carreau fits are of order 2%. We note that these errors can be further reduced by a factor of 10 by fitting the data to the Carreau–Yasuda equation that uses an additional adjustable parameter [5]. There is a gradual decrease in the exponent n of the power law as P increases; n decreases from about 0.46 at 0.1 MPa to 0.2 at 200 MPa.

Figure 2b shows the viscosities η corresponding to the stress data in Fig. 2a in a log–log plot. Data for $P < 300$ MPa in Fig. 2b are fit to the Carreau equation with values of n that decrease with increasing pressure. Data at higher

pressures also have apparently linear regions in Fig. 2b with a slope near -1 that corresponds to n near zero. However, as noted already, the data at these pressures are better fit by the Eyring equation where the high frequency limit, $\eta \sim \log \dot{\gamma} / \dot{\gamma}$, is equivalent to a power law in the limit of n approaching zero. Eyring theory describes the decrease in viscosity by six orders of magnitude relative to the Newtonian viscosity η_N . Although there are no non-equilibrium experimental data for $P \leq 500$ MPa, simulation data are consistent with the experimental η_N values [9, 11, 31]. Eyring theory also provides a reasonable fit for systems prepared under different T and P as long as $\eta_N \gtrsim 1$ Pa·s. A collapse of reduced viscosity η/η_N vs rate scaled by $\dot{\gamma}_E$ is found for all systems with $\eta_N \gtrsim 1$ Pa·s. Over almost 30 decades, simulation and experimental data lie on the analytic solution of the Eyring model.

Shear thinning in the Eyring model originates from a mechanism that only involves a change in the rate of rearrangements through activated motion over fixed barriers. In contrast, one commonly invoked mechanism for power-law shear thinning described by the Carreau model is a change in some order parameter. For squalane, the simplest type of order is the alignment of molecules along the flow direction [23, 55, 56]. This may lower the viscosity by decreasing the rate of collisions between molecules and the associated momentum transfer and dissipation. To study changes in molecular order under shear, Jadhao and Robbins [9] extracted the orientational tensor using NEMD simulation data:

$$S_{\alpha\beta} = \frac{3}{2} \left\langle \frac{1}{N_m} \sum_{i=1}^{N_m} u_{i\alpha} u_{i\beta} - \frac{1}{3} \delta_{\alpha\beta} \right\rangle, \quad (3)$$

where α and β are Cartesian coordinates, $\hat{u}_i$ is the unit vector in the direction between the end atoms of the i th squalane molecule, and N_m is the number of molecules. The diagonal elements S_{xx} , S_{yy} , and S_{zz} measure the degree of molecular alignment in the direction of the flow field, velocity gradient, and vorticity, respectively. A value of 0 indicates no preference to align along the axis, while values of 1 and -0.5 imply perfect alignment along or perpendicular to the axis.

Figure 3 shows the variation of S_{xx} and S_{yy} with rate $\dot{\gamma}$ for squalane sheared at $T = 293$ K and pressures P indicated in the legend. In equilibrium, the system must become isotropic, implying $S_{xx} = S_{yy} = 0$. This limit is approached for the lowest P and $\dot{\gamma}$ in Fig. 3. With increasing $\dot{\gamma}$, molecules become more aligned along the flow direction. S_{xx} and S_{yy} rise rapidly and then saturate for all P . These changes occur over about two decades in rate. The maximum degree of alignment increases only slightly with increasing P . By correlating the rate dependence of orientational order and the reduced viscosity, it is found that the alignment saturates when viscosity η/η_N drops by

Fig. 3 Rate dependence of orientational order components S_{xx} and S_{yy} at the pressures P indicated in the legend. The results shown are obtained in [9]. Error bars are of order ± 0.03 except at the lowest rates and highest P . Changes in alignment are confined to about two decades in rate

only a factor of ≈ 3 . The viscosity then continues to drop by many decades with little change in alignment.

3 Machine Learning Techniques for Analyzing Rheological Behavior

The link between rheological properties and changes in the molecular order in squalane can be examined more closely by visualizing the information encoded in all the components of the orientation tensors associated with all the atom pairs. We now describe the data preparation processes and the ML techniques used to enable the rapid visualization of this high-dimensional information.

3.1 Simulation Data Collection and Preparation

NEMD simulation produces trajectory data consisting of 3D positions of $N_m \times 30 = 3750$ atoms, where $N_m = 125$ is the number of squalane molecules (each molecule has 30 united atoms). We collect this trajectory data from simulations at $T = 293$ K for six pressures $P = 0.1, 100, 400, 636, 875, 955$ MPa and rates $\dot{\gamma} \in 10^6 - 10^{10} \text{ s}^{-1}$. For all rates except 10^6 s^{-1} , this information is collected at a strain interval of 0.01. That is, the data collection frequency f changes with $\dot{\gamma}$; for $\dot{\gamma} = 10^{10} \text{ s}^{-1}$, $f = 1000$ steps and for $\dot{\gamma} = 10^7 \text{ s}^{-1}$, $f = 10^6$. As simulations at $\dot{\gamma} = 10^6 \text{ s}^{-1}$ are relatively slower, the data are collected for every 0.001 strain. In some higher $\dot{\gamma}$ cases, new simulations are run to produce trajectory data with the needed granularity of strain. Otherwise, trajectory data generated from past simulations performed by Jadhao and

Robbins are utilized [9, 31]. For the purposes of dimension reduction, the last 1000 frames of the trajectory data associated with the steady-state region are used to compute the pair orientation tensors (see below). For each system (specified by pressure and rate), the size of this dataset is ~ 4 GB. The complete dataset size is ~ 200 GB.

The trajectory data is used to extract the orientation tensor $S_{\alpha\beta}^p$ for each atom pair p belonging to the same molecule:

$$S_{\alpha\beta}^p = \frac{3}{2} \left\langle \frac{1}{N_m} \sum_{i=1}^{N_m} r_{i\alpha}^p r_{i\beta}^p - \frac{1}{3} \delta_{\alpha\beta} \right\rangle, \quad (4)$$

where α and β are Cartesian coordinates, and r_i^p is the unit distance vector between the atoms associated with pair p of the i th squalane molecule. The diagonal elements S_{xx}^p , S_{yy}^p , and S_{zz}^p of the pair orientational tensor measure the degree of alignment of the atoms associated with pair p in the direction of the flow field, velocity gradient and vorticity, respectively. Each element of the tensor $S_{\alpha\beta}^p$ represents an ensemble average of the distance between the atoms associated with the pair p . The ensemble average is performed over the number of squalane molecules and over the selected time window (Fig. 4). Note that $S_{\alpha\beta}^p$ is equivalent to $S_{\alpha\beta}$ of Eq. 3 for the end atom pair (1–30) of the squalane molecule.

$S_{\alpha\beta}^p$ is a 3×3 symmetric matrix, that is $S_{\alpha\beta}^p = S_{\beta\alpha}^p$. This means six of the nine components are non-trivial and fully specify $S_{\alpha\beta}^p$. Further, there are a total of $N_p = 435$ distinct pairs associated with a squalane molecule. Figure 4 shows a sketch of the averaging process used to obtain the orientation tensor for all the 435 pairs. Given the six non-trivial

Fig. 4 Ensemble averaging to get the orientation tensors for 435 pairs using the trajectory data generated by non-equilibrium molecular dynamics simulations. Horizontal dots indicate simulation time and vertical dots indicate the number of molecules. Multiple copies indicate the number of pairs

components of $S_{\alpha\beta}^p$ and $N_p = 435$ distinct pairs, the orientation information is represented with $6 \times 435 = 2610$ dimensions. We did not apply a min-max normalization to the high-dimensional data in order to preserve the relative difference between the separate components of the orientation tensor in the reduced dimension.

3.2 Dimension Reduction Methods

We now describe the ML techniques that enable the visualization of the 435×6 dimensional information associated with the pair orientation tensors in reduced dimensions. We experiment with two dimension reduction techniques: principal component analysis and t-distributed stochastic neighbor embedding, and reduce the data of N dimensions to 2. The $N = 6$ to 2 dimension reduction reduces the components of $S_{\alpha\beta}^p$ for each pair p , and the $N = 2610$ to 2 dimension reduction reduces the information encoded in all components of all pairs for a given pressure and rate. Our code and the data utilized for dimension reduction tasks are available on GitHub.¹

3.2.1 Principal Component Analysis (PCA)

PCA is an unsupervised linear transformation technique used for feature extraction and dimension reduction [57, 58]. It enables the visualization of a dataset by performing a linear mapping of the data to a lower dimensional space so that the variance of the data in the reduced dimensional space is maximized. The linear transformation chooses a new coordinate system for the dataset such that the greatest

variance by any projection of the dataset comes to lie on the first axis, then called the first principal component (or PC1), which accounts for as much of the variability in the data as possible. The second greatest variance lies on the second axis (PC2) under the constraint that it is orthogonal to the preceding component. Large pairwise distances are preserved after the original data are projected on to the two principal component axes.

The implementation involves finding the eigenvalues and eigenvectors of the $N \times N$ dimensional covariance matrix associated with the N -dimensional input dataset. The covariance matrix is symmetric and its diagonal elements are the variances of N dimensions. The eigenvectors of the covariance matrix with the largest eigenvalues are identified as unit vectors in the N dimensional space associated with the principal components PC1 and PC2. These components describe the largest variations in the projected high-dimensional data in reduced dimensions. PC1 accounts for the maximum variation in the projected data, and PC2 represents the axis orthogonal to PC1 that is associated with the next largest variation. The PCA utility provided in the scikit-learn package is used to perform the PCA [59].

3.2.2 t-Distributed Stochastic Neighbor Embedding (t-SNE)

t-SNE is a popular non-linear dimension reduction technique which captures non-linear relationships between the elements of the dataset that are not captured by linear methods such as PCA [60]. However, the interpretability of the patterns observed in data reduced via t-SNE is limited compared to PCA. In t-SNE, the local structures associated with the data are preserved while converting from higher dimensions to lower dimensions. t-SNE constructs a probability distribution over pairs of high-dimensional objects such that similar objects have a high probability of being picked, while dissimilar objects have a very low probability of being picked. It defines a similar probability distribution over objects in low dimensions. Finally, it minimizes the Kullback-Leibler divergence between the two distributions with respect to the locations of the objects in respective high and low dimensional spaces.

We use the t-SNE utility provided in the scikit-learn package [59]. The key parameters from an implementation standpoint are: perplexity and n_iter. Other optimization parameters are set to their default values provided in the scikit-learn library. Perplexity is related to the number of nearest neighbors. We tuned perplexity over a range of 20 to 60 for the different dimension reduction tasks. n_iter represents the maximum number of iterations scikit-learn will perform for a given dimension reduction task. We find that the maximum n_iter value required across all tasks is 3000 via a grid search for the n_iter parameter. So, n_iter is set to 3000 across all studies. We note that the optimization

¹ Available at <https://github.com/softmaterialsmlflow/>.

Fig. 5 Behavior of 435 atom pairs associated with a squalane molecule illustrated via dimension reduction of orientation tensors using PCA for a single strain rate $\dot{\gamma}$ under ambient pressure $P = 0.1$ MPa at $T = 293$ K. Each point represents one pair. Shear rate labels (“0”, “ 10^9 ”) show the mean point with respect to all pairs. Side, end, backbone, and end-to-end pairs are labeled with pair numbers in blue,

lime green, olive green, and magenta colors, respectively. The values in the brackets on the PC1 and PC2 axis labels are the variance ratios which signify the amount of variance captured along the respective axis. **a** Unsheared squalane ($\dot{\gamma} = 0$). **b** Squalane sheared at $\dot{\gamma} = 10^9 \text{ s}^{-1}$ (Color figure online)

involved in dimension reduction uses an early stopping criterion based on the number of iterations needed for convergence, regardless of the `n_iter` value we set. The number of iterations required to converge a dimension reduction task depends on the complexity of the task. Here, convergence indicates that the visualization in the reduced dimensions is no longer changing with increasing the number of iterations.

4 Results

We now utilize the information encoded in all 435 pairs as well as all the six non-trivial components of the pair orientation tensors to make a complete assessment of the correlation between strain rate, pressure, and molecular alignment in squalane via dimension reduction methods. We begin by showing the results for the dimension reduction of the 6-dimensional dataset formed by the components of $S_{\alpha\beta}^p$ to 2 dimensions for each pair p .

4.1 Rheological Response in Ambient Conditions

4.1.1 Dimension Reduction using PCA at a Single Rate

We first discuss the results obtained at a single strain rate using PCA for squalane under ambient conditions of $P = 0.1$ MPa. It is useful to recall that during the data pre-processing step for PCA, the data for each component of the pair orientation tensor (e.g., S_{xx}^p) are centered using the mean of the components across all 435 pairs (e.g., $\sum_{p=1}^{435} S_{xx}^p$), and no normalization is applied to the high-dimensional data.

Figure 5a shows a scatter plot of the six dimensional $S_{\alpha\beta}^p$ components reduced to two dimensions using PCA for 435 pairs associated with the unsheared squalane system. We label this system with “0” strain rate, noting that this result corresponds to an equilibrium NVT simulation of squalane. Each point in Fig. 5a represents one orientation tensor $S_{\alpha\beta}^p$ for an atom pair p . Almost all pairs have ≈ 0 values (scores) along the two principal component axes, PC1 and PC2. The scores for PC1 and PC2 are similar and these components capture a similar amount of variance in the tensor components (72% and 28% by PC1 and PC2, respectively). The pairs are also nearly symmetrically distributed around the rate label 0 which represents the center of all 435 pairs relative to both PC1 and PC2.

Squalane has 4 side atom pairs, 4 end atom pairs, 4 end-to-end atom pairs, and 15 physically connected backbone pairs. On the plot, we highlight these distinct groups of atom pairs using the following sets: 4 side atom pairs (7–8, 12–13, 18–19, 23–24), 4 end atom pairs (1–3, 2–3, 28–29, 28–30), 4 randomly chosen physically connected backbone atom pairs (3–4, 11–12, 20–21, 27–28), and 4 end-to-end atom pairs (1–30, 1–29, 2–30, 2–29). These are marked with the atom indices associated with the pair in different colors as noted in the caption. While some of the side atom pairs (12–13, 7–8) appear to contribute the largest variation in the data, no clear and significant separation between the different groups of atom pairs is observed. Overall, PCA results suggest that under ambient conditions, squalane acts like an isotropic liquid at zero strain rate. We note that the deviations in the PC1 and PC2 scores from 0 observed in the data for this case are due to finite size effects. These scores provide a reference

to interpret the effects of shearing squalane with non-zero strain rates and under higher pressures.

Figure 5b shows the PCA results for squalane sheared at a strain rate of 10^9 s^{-1} , which is chosen as a representative case for the sheared systems. These results are in clear contrast with the zero strain rate case. Most of the 435 pairs now have scores significantly different from 0 along both PC1 and PC2. The PC1 scores are significantly larger ($\approx 5\times$) than PC2 scores, and PC1 captures 96% of the variation in the tensor components associated with the pairs. The average pair behavior coincides with the center ($\text{PC1} = 0, \text{PC2} = 0$) as evidenced by the “ 10^9 ” rate label, and is similar to the zero strain rate case. However, the distribution of the pairs in the 2D scatter plot is clearly asymmetric and inhomogeneous, indicating that many pairs contribute differently to the variations in the orientation tensor components.

We now examine these differences using the 4 groups of atom pairs defined above. Unlike the unsheared system, PCA clearly groups these four sets of atom pairs into well-separated clusters for squalane sheared at 10^9 s^{-1} . The side atom pairs (blue) account for the largest variation in the pair orientation tensor components $S_{\alpha\beta}^p$ along PC1, followed by the end atom pairs (lime green) and the set of backbone atom pairs (olive green). The end-to-end pairs also contribute significantly albeit in a contrasting role as indicated by the negative sign associated with their PC1 scores. The large density of points near the end-to-end pairs corresponds to pairs with atoms separated by at least 25 atoms. These pairs play a similar role as the end-to-end atom pairs in explaining the variation in the data.

The large variations across the $S_{\alpha\beta}^p$ components exhibited by the side (and end) atom pairs can be attributed to their atoms being relatively less constrained in responding to thermal forces which can induce changes in all 6 $S_{\alpha\beta}^p$ components. On the other hand, the end-to-end atom pairs respond strongly to shear-driven alignment which couples most significantly to the flow and gradient directions and likely induces large changes in the $S_{xx}^p, S_{xy}^p, S_{yy}^p$ components. The PCA scatter plot establishes that these two distinct groups of atom pairs contribute in contrasting roles.

4.1.2 Dimension Reduction of Pair Orientation Tensors for Different Rates

We now use PCA to reduce the 6-dimensional $S_{\alpha\beta}^p$ tensors to 2 dimensions for each of the 435 pairs among all 8 strain rates $\dot{\gamma} : 0, 10^7, 2 \times 10^7, 10^8, 2 \times 10^8, 10^9, 2 \times 10^9, 10^{10} \text{ s}^{-1}$. Our aim is for PCA to furnish the principal components that enable the examination of the competing contributions of side atom pairs vs end-to-end atom pairs to the variations in the $S_{\alpha\beta}^p$ components across all pairs and all strain rates. We also expect PCA to reveal any significant changes in the flow

Fig. 6 Behavior of atom pairs associated with a squalane molecule illustrated via dimension reduction of orientation tensors using PCA for 8 strain rates $\dot{\gamma} = 0, 10^7, 2 \times 10^7, 10^8, 2 \times 10^8, 10^9, 2 \times 10^9$, and 10^{10} s^{-1} under ambient conditions. Each point represents one pair orientation tensor. Each rate is represented by 435 pairs. Different symbols of points represent different rates. Shear rate label with the same color marks the center of the data with respect to all the pairs associated with that rate. Side, backbone, and end-to-end pairs are labeled with pair numbers in blue, olive green, and magenta colors, respectively. The values in the brackets on the PC1 and PC2 axis labels are the variance ratios which signify the amount of variance captured along the respective axis (Color figure online)

behavior vs rate and provide a complete assessment of the extent of shear thinning in squalane resulting from changes in molecular alignment.

Figure 6 shows the visualization of the dimension reduced $S_{\alpha\beta}^p$ tensors on a 2D scatter plot using PCA. Each point represents an orientation tensor $S_{\alpha\beta}^p$. Different symbols correspond to different strain rates and are manually assigned without invoking any clustering algorithm. The pair orientation tensor data for each rate are represented by 435 points. Labels for each strain rate are placed at the center of the data of that rate to facilitate assessing the separation between different clusters. We also mark the atom indices associated with the side atom pairs, randomly chosen backbone atom pairs, and end-to-end atom pairs in different colors as noted in the caption. The index labels for the end atom pairs are not shown for the sake of clarity.

We first note that PC1 scores are significantly larger ($5\times$) compared to PC2 scores. The values 0.95 and 0.03 on the axis labels PC1 and PC2 indicate variance ratios that signify the amount of variance in the data captured along the respective axis. Using the variance ratios, we

find that PC1 captures 97% of the variance in the data. Atom pairs belonging to high rates ($\dot{\gamma} \gtrsim 10^9 \text{ s}^{-1}$) have high PC1 scores. The end-to-end atom pairs contribute the largest variation in the $S_{\alpha\beta}^p$ components for a given high rate. Examining the rate labels shows that the clusters of $S_{\alpha\beta}^p$ data for different rates are well separated from each other and organized along PC1 by strain rate increasing from left to right. This indicates that changes in shear strain rate strongly correlate with the variations in $S_{\alpha\beta}^p$ components. The $S_{\alpha\beta}^p$ data for end-to-end atom pairs at different rates is grouped into well-separated clusters. The associated PC1 scores range from ~ -0.1 to $\sim +0.4$ as the rates shift from 10^7 to 10^{10} s^{-1} . This large evolution in the orientation tensors for end-to-end atom pairs in 2 dimensions likely originates from significant changes in S_{xx}^p , S_{yy}^p , and S_{xy}^p in the 6-dimensional space because these components couple most strongly to shear. This signals a strong correlation between changes in molecular alignment and the shear thinning of squalane over this range of rates.

Figure 6 shows that the side atom pairs have high negative PC1 scores, indicating that they contribute to the variation in $S_{\alpha\beta}^p$ components in a contrasting role to the end-to-end atom pairs. These roles are reversed in comparison to the system sheared at a single rate of 10^9 s^{-1} ; Fig. 5b. Further, the side atom pairs for all rates are grouped together, indicating that the changes in strain rate minimally affect the $S_{\alpha\beta}^p$ components associated with these pairs.

PCA shows that the pair orientation tensor datasets for rates $\dot{\gamma} \leq 2 \times 10^8 \text{ s}^{-1}$ are grouped in a region (with PC1 scores $\lesssim 0$) that is well separated from the datasets for $\dot{\gamma} > 2 \times 10^8 \text{ s}^{-1}$. The separation of clusters for $\dot{\gamma} \leq 2 \times 10^8 \text{ s}^{-1}$ from the $S_{\alpha\beta}^p$ datasets for $\dot{\gamma} > 2 \times 10^8 \text{ s}^{-1}$ suggests a significant change in the pair orientation tensors as the strain rate increases beyond $2 \times 10^8 \text{ s}^{-1}$. This shift can be associated with the change in the flow behavior seen in the viscosity vs rate data in Fig. 2b from Newtonian flow to non-Newtonian shear thinning with increasing rate.

The observation of the crossover between the Newtonian and non-Newtonian flow regimes is made more precise and clear via dimension reduction using t-SNE. Figure 7 shows the dimension reduced $S_{\alpha\beta}^p$ tensors using t-SNE for the same set of physical systems analyzed using PCA. Different strain rates are represented with the same color scheme and marker style as in PCA. Following standard practices, we have not shown the abscissae and ordinates explicitly in the plot because t-SNE preserves the differences between samples in the high-dimensional space in transforming the information into two dimensions to produce abscissae and ordinates whose absolute values are not comparable and meaningful on their own [60, 61]. Each point corresponds to one pair orientation tensor $S_{\alpha\beta}^p$. t-SNE groups pairs associated with

Fig. 7 Behavior of atom pairs associated with a squalane molecule illustrated via dimension reduction of orientation tensors using t-SNE for 8 strain rates $\dot{\gamma} = 0, 10^7, 2 \times 10^7, 10^8, 2 \times 10^8, 10^9, 2 \times 10^9$, and 10^{10} s^{-1} under ambient conditions. Each point represents one pair orientation tensor. Each rate is represented by 435 pairs. Different symbols represent different rates. Shear rate label with the same color marks the center of the data with respect to all the pairs associated with that rate. Side, backbone, and end-to-end pairs are labeled with pair numbers in blue, olive green, and magenta colors, respectively (Color figure online)

each rate into well-separated clusters for almost all pairs, indicating that the $S_{\alpha\beta}^p$ components are more clearly separable based on rates compared to what PCA revealed. A clear gap is observed between the clusters for $\dot{\gamma} < 2 \times 10^8 \text{ s}^{-1}$ (top left) and those for $\dot{\gamma} \geq 2 \times 10^8 \text{ s}^{-1}$ (bottom right). Unlike PCA, t-SNE groups the $S_{\alpha\beta}^p$ data associated with 435 pairs at $2 \times 10^8 \text{ s}^{-1}$ closer to the $S_{\alpha\beta}^p$ data at higher rates. The clear separation between pair orientation tensor data for $\dot{\gamma} \geq 2 \times 10^8 \text{ s}^{-1}$ and $\dot{\gamma} < 2 \times 10^8 \text{ s}^{-1}$, and the prediction of a crossover rate between 1×10^8 and $2 \times 10^8 \text{ s}^{-1}$ is in agreement with the Newtonian to non-Newtonian transition in the shear flow response observed in the viscosity vs rate plot shown in Fig. 2b.

t-SNE also confirms the PCA results on the contrasting roles of the end-to-end atom pairs and the side atom pairs in governing the evolution in $S_{\alpha\beta}^p$ components with rate. The side atom pairs are among the least well-separated set of atom pairs. On the other hand, the end-to-end atom pairs best classify the $S_{\alpha\beta}^p$ data for different rates as they are among the most well-separated set of atom pairs. This suggests that the important differences in the high-dimensional dataset are

Fig. 8 Behavior of atom pairs associated with a squalane molecule illustrated via dimension reduction of orientation tensors using PCA for six different pressures: **a** 0.1 MPa, **b** 100 MPa, **c** 400 MPa, **d** 636 MPa, **e** 875 MPa, and **f** 955 MPa and associated strain rates $\dot{\gamma} \in (10^6, 10^{10}) \text{ s}^{-1}$ described in the text. Each point represents one pair orientation tensor. Different symbols represent different rates as shown in the legend panel at the top. Symbols are kept uniform across different subplots (e.g., green diamond represents 435 pairs at

$\dot{\gamma} = 10^7 \text{ s}^{-1}$ across all six pressures.) Shear rate label with the same color marks the center of the data with respect to all the pairs associated with that rate. Side, backbone, and end-to-end pairs are labeled with pair numbers in blue, olive green, and magenta colors, respectively. The values in the brackets on the PC1 and PC2 axis labels are the variance ratios which signify the amount of variance captured along the respective axis (Color figure online)

originating from the $S_{\alpha\beta}^p$ features associated with the end-to-end atom pairs.

Overall, the dimension reduction of pair orientation tensor data involving all 435 pairs and 6 $S_{\alpha\beta}^p$ components via PCA and t-SNE shows clear evidence of a strong correlation between shear strain rate and the evolution of $S_{\alpha\beta}^p$. Changes in $S_{\alpha\beta}^p$ are dominated by the end-to-end atom pairs when data across all systems sheared at different strain rates is analyzed. A clear distinct grouping of datasets for rates $\dot{\gamma} \geq 2 \times 10^8 \text{ s}^{-1}$ signals a dramatic change in the flow behavior at a rate between 1×10^8 and $2 \times 10^8 \text{ s}^{-1}$. This result can be utilized to determine the characteristic rate for the onset of shear thinning in the Carreau model ($\dot{\gamma}_0$).

4.2 Rheological Response Under Different Pressures

We now present the results for the dimension reduction of the pair orientation tensor data associated with the shear flow of squalane under different pressures. Figure 8 shows six scatter plots representing the PCA results for six different pressures $P = 0.1, 100, 400, 636, 875, 955 \text{ MPa}$. Results for the ambient pressure $P = 0.1 \text{ MPa}$ are re-plotted in Fig. 8a and serve as a reference to assess changes in $S_{\alpha\beta}^p$ with increasing P . For $P = 100 \text{ MPa}$, results are shown for the same set of 8 strain rates as used in the system under ambient conditions: $0, 10^7, 2 \times 10^7, 10^8, 2 \times 10^8, 10^9, 2 \times 10^9, 10^{10} \text{ s}^{-1}$. For $P = 400 \text{ MPa}$, in addition to these rates, results are shown for 10^6 s^{-1} . For $P = 636, 875, 955 \text{ MPa}$, results are shown for 2 additional rates: 10^6 and $2 \times 10^6 \text{ s}^{-1}$. Other details are the same as noted for the system under ambient conditions. We note that the PCA

plots and the associated scores remain unchanged by excluding the zero strain rate system from the dimension reduction tasks for the different pressures reported in Fig. 8. We have included the dimension reduction results for the zero rate case as a reference for each pressure to ease the assessment of changes in the flow response of squalane with increasing rate (for example, by examining the separation of the different clusters from the “0” labeled cluster). For all pressures, PC1 scores are significantly larger ($\approx 5\times$) than PC2 scores. PC1 captures 96–98% of the variance in the data for the pressures studied.

For 100 MPa, the rate labels for the clusters of $S_{\alpha\beta}^p$ data for many rates are well separated from each other. The clusters are organized by strain rate along PC1, however, in contrast to the ambient case, the PC1 scores now increase with *decreasing* rate. This switch can be attributed to the large contributions of the side atom pairs to the variations in the pair orientation tensor components, as reflected by their large positive PC1 scores, which surpass those of the end-to-end atom pairs that receive large negative PC1 scores. The end-to-end atom pairs provide the clearest separation of $S_{\alpha\beta}^p$ data associated with different rates, although the separation is worse for the highest rates ($10^9, 10^{10} \text{ s}^{-1}$) compared to the ambient case. The associated PC1 scores range from ~ -0.3 to $\sim +0.15$ as the rates shift from 10^7 to 10^{10} s^{-1} . These scores vary over a narrower range compared to the ambient case suggesting that the evolution in $S_{\alpha\beta}^p$ for the end-to-end atom pairs and related changes in molecular order saturate relatively quickly. Despite the overall greater degree of overlap between clusters of $S_{\alpha\beta}^p$ data for different rates compared to the ambient case, there is a correlation between changes in strain rate, molecular alignment, and shear thinning of squalane over the examined range of rates for 100 MPa.

PCA shows that the pair orientation tensor datasets for rates $\dot{\gamma} \leq 2 \times 10^7 \text{ s}^{-1}$ are grouped in a region (with PC1 scores $\gtrsim 0$) that is well separated from the datasets for $\dot{\gamma} > 2 \times 10^7 \text{ s}^{-1}$. This distinct grouping of clusters of $S_{\alpha\beta}^p$ datasets for $\dot{\gamma} > 2 \times 10^7 \text{ s}^{-1}$ is similar to that observed in the ambient case (albeit at a different crossover rate), and signals a similar significant change in the pair orientation tensors and flow behavior as the strain rate increases beyond $2 \times 10^7 \text{ s}^{-1}$. As before, a more precise assessment of this crossover is obtained via the dimension reduction using t-SNE shown later.

PCA results for 400 MPa show significant differences from the results at lower pressures $P = 0.1, 100 \text{ MPa}$. First, the labels for the clusters of $S_{\alpha\beta}^p$ data associated with different strain rates show no discernible ordering of strain rates along PC1 (or PC2). These rate labels are also not clearly separable for most rates except for the zero strain rate system. The $S_{\alpha\beta}^p$ data for $1 \times 10^6 \text{ s}^{-1}$ are separated from the data

for the higher strain rates but only marginally as indicated by its rate label. In stark contrast with results at lower pressures, the end-to-end atom pairs do not clearly separate $S_{\alpha\beta}^p$ tensors associated with different strain rates for 400 MPa. The associated PC1 scores range from ~ -0.2 to ~ -0.05 as the rates shift from 10^6 to 10^{10} s^{-1} . These scores vary over a much narrower range compared to the results for $P = 0.1, 100 \text{ MPa}$ systems, signaling that the evolution in $S_{\alpha\beta}^p$ vs rate for the end-to-end atom pairs and related changes in molecular order saturate relatively quickly. These observations imply that the extent of shear thinning of squalane at 400 MPa arising from shear-driven molecular alignment is limited.

The side atom pairs contribute the largest variation in the pair orientation tensor components along PC1, easily surpassing the contributions of the end-to-end atom pairs. These atom pairs (identified by blue atom indices) separate out into a cluster and are not separable by rate. The next largest variation in $S_{\alpha\beta}^p$ components is contributed by the end atom pairs (indices not labeled) and by the atom pairs associated with the system at zero strain rate. The closely spaced clusters associated with these three sets of atom pairs suggest the presence of a common mechanism. The dominant contribution to shear thinning of squalane as it is pushed toward the glassy regime at 400 MPa may originate from the thermally activated rearrangements of the side and the end atom pairs. New investigations utilizing different types of high-dimensional data are required to make progress toward a more precise diagnosis.

PCA results for higher pressures of $P = 636, 875, 955 \text{ MPa}$ share many similarities with the results for 400 MPa. The rate labels are stacked almost on top of each other. There is no evidence of any ordering of strain rates along PC1 or PC2. Except for the unsheared system, there is no separation of $S_{\alpha\beta}^p$ data by strain rate and no clear grouping of clusters of $S_{\alpha\beta}^p$ datasets for different sets of low and high rates (as seen for $P = 0.1, 100 \text{ MPa}$). The latter observation confirms that the flow behavior for the probed rates under these pressures is squarely within the shear thinning region; Fig. 2b. The ability of the end-to-end atom pairs to separate the data at different strain rates worsens as P is increased. The associated PC1 scores range from ~ -0.2 to ~ -0.05 , similar to 400 MPa, as the rates shift from 10^6 to 10^{10} s^{-1} . These scores vary over a much narrower range compared to the results for $P = 0.1, 100 \text{ MPa}$ systems, signaling that the evolution in $S_{\alpha\beta}^p$ vs rate for the end-to-end atom pairs and related changes in molecular order saturate relatively quickly. These observations imply that the contribution to shear thinning of squalane at high P due to shear-driven molecular alignment is small and bounded.

Fig. 9 Behavior of atom pairs associated with a squalane molecule illustrated via dimension reduction of orientation tensors using t-SNE for six different pressures **a** 0.1 MPa, **b** 100 MPa, **c** 400 MPa, **d** 636 MPa, **e** 875 MPa, and **f** 955 MPa and associated strain rates $\dot{\gamma} \in (10^6, 10^{10}) \text{ s}^{-1}$ described in the text. Each point represents one pair orientation tensor. Different symbols represent different rates as shown in the legend panel at the top. Symbols are kept uniform

The side atom and end atom pairs dominate the contribution to the variation in the pair orientation tensor components along PC1. Their PC1 scores slightly increase with pressure. Similar to the case of 400 MPa, the clusters formed by these groups of atom pairs are close to the cluster associated with $S_{\alpha\beta}^p$ data for all pairs belonging to the unsheared system. As squalane is pushed deep into the glassy flow regime under these high pressures of $P = 636, 875, 955 \text{ MPa}$, we expect the common mechanism of thermally activated rearrangements of the side and end atoms to dominate the contribution to the decrease in viscosity with rate.

We now discuss the results of dimension reduction for different P using t-SNE. Figure 9 shows the visualization of the dimension reduced $S_{\alpha\beta}^p$ tensors (from 6 dimensions to 2 dimensions) using t-SNE for the same set of physical systems analyzed using PCA in Fig. 8. For each pressure, different strain rates are represented with the same symbol scheme as in Fig. 8; other details are also the same as in the PCA plots. Results for the ambient pressure $P = 0.1 \text{ MPa}$ are

across different subplots (e.g., green diamond represents 435 pairs at $\dot{\gamma} = 10^7 \text{ s}^{-1}$ across all six pressures.) Shear rate label with the same color marks the center of the data with respect to all the pairs associated with that rate. Side, backbone, and end-to-end pairs are labeled with pair numbers in blue, olive green, and magenta colors, respectively (Color figure online)

re-plotted in Fig. 9a to serve as a reference to assess changes in $S_{\alpha\beta}^p$ with increasing P .

Figure 9b shows t-SNE results for $P = 100 \text{ MPa}$, which share many similarities with the results for the ambient case. t-SNE groups atom pairs comprising the $S_{\alpha\beta}^p$ data associated with different strain rates into well-separated clusters for a majority of pairs, signaling that the 6-dimensional $S_{\alpha\beta}^p$ tensor components are clearly separable based on rates. Similar to the 0.1 MPa system, t-SNE shows that the end-to-end atom pairs best classify the $S_{\alpha\beta}^p$ data for different rates as they are among the most well-separated set of atom pairs occupying the farthest regions in the 2D scatter plot. t-SNE confirms the PCA results on the contrasting roles of the end-to-end atom pairs and the side atom pairs in governing the evolution in $S_{\alpha\beta}^p$ components with rate. The side atom pairs are among the least well-separated set of atom pairs.

There are few differences between the t-SNE results for the ambient case and $P = 100 \text{ MPa}$ system. The side atom pairs and the backbone atom pairs organize in larger

numbers into more clearly separable and dense clusters for 100 MPa. An increase in the overlap of pairs (e.g., atom pairs separated by at least 25 atoms) associated with rates $\dot{\gamma} \geq 10^8 \text{ s}^{-1}$ is observed for 100 MPa compared to the ambient case.

t-SNE results for 100 MPa show a clear gap between the clusters for $\dot{\gamma} \leq 2 \times 10^7 \text{ s}^{-1}$ and those for $\dot{\gamma} > 2 \times 10^7 \text{ s}^{-1}$. This distinct grouping of pair orientation tensors and the prediction of a crossover rate between 2×10^7 and $1 \times 10^8 \text{ s}^{-1}$ are in agreement with the Newtonian to non-Newtonian transition in the shear flow response observed in the viscosity vs rate plot shown in Fig. 2b. This crossover rate is lower than that for the ambient conditions.

Figure 9c shows the t-SNE results for $P = 400$ MPa which show important differences from the results at lower pressures. The atom pairs belonging to the $S_{\alpha\beta}^p$ data associated with different strain rates are not easily separable for most rates. Only the atom pairs for the 10^6 s^{-1} rate and the unsheared system are separately grouped together into clusters well separated from the atom pairs that belong to higher strain rates. Unlike PCA results, the end-to-end atom pairs are able to separate the different strain rates but they are no longer the most well-separated and spread out pairs as in the case of $P = 0.1, 100$ MPa systems; Fig. 9a and b. The distance between clusters of the end-to-end atom pairs for higher rates is substantially reduced, and the pairs appear to cluster on one side of the 2D scatter plot.

Similar to PCA, t-SNE groups the side atom pairs (blue indices) and the end atom pairs for all strain rates into separate clusters. The clusters formed by these groups of atom pairs are close to the $S_{\alpha\beta}^p$ data cluster for all pairs associated with the unsheared system. t-SNE results for 400 MPa also show a relatively more clear separation between the data at higher rates and the $S_{\alpha\beta}^p$ cluster at 10^6 s^{-1} . Including the pair orientation tensor data (not shown) from $\dot{\gamma} = 3 \times 10^5 \text{ s}^{-1}$ in the dimension reduction task for 400 MPa, t-SNE predicts a crossover rate to Newtonian flow occurs in the vicinity of 10^6 s^{-1} , in agreement with Fig. 2b.

t-SNE results for the higher pressures of $P = 636, 875, 955$ MPa share many similarities with the corresponding PCA results. The rate labels for $S_{\alpha\beta}^p$ data associated with different decades of rate are nearly stacked on top of each other, with the separation getting worse as P increases. Only the data for the unsheared system are clearly separated by strain rate. The end-to-end atom pairs separate the data for different strain rates, but the distance between these clusters is further reduced compared to the 400 MPa system. These pairs are organized on one side of the 2D scatter plot (e.g., bottom left for $P = 875, 955$ MPa) and are substantially less spread out relative to the large separations observed for lower $P = 0.1, 100$ MPa systems.

Similar to PCA, t-SNE groups the side atom pairs (blue indices) and the end atom pairs (not labeled) for all strain rates into separate clusters. The clusters formed by these groups of atom pairs are close to the $S_{\alpha\beta}^p$ data cluster for all pairs associated with the unsheared system. The separation between these different clusters decreases with increasing pressure. This further supports the existence of a common mechanism governing the dynamics of atom pairs associated with these groups. We note that t-SNE groups the backbone atom pairs for all non-zero rates together into a cluster well separated from the side and end atom pairs. Furthermore, the large mix of atom pairs in the middle of the scatter plots for these high pressures corresponds to pairs separated with the number of atoms between 3 and 25.

4.3 Single Dimension Reduction for All Systems Using PCA

We now discuss the PCA results for the dimension reduction of the high-dimensional pair orientation tensor data associated with the shear flow of squalane for all non-zero strain rate values and all pressures using a single dimension reduction plot. The principal components PC1 and PC2 furnished by PCA employ the information from 49 systems characterized with different rates and pressures.

Figure 10a shows the reduction of the 6-dimensional $S_{\alpha\beta}^p$ tensors to 2 dimensions for pressures $P = 0.1, 100, 400, 636, 875, 955$ MPa and associated values of strain rates described in Section 4.2. For each pressure, at a given rate, 435 pairs comprise the scatter points shown on the 2D plot. The legend block on the right shows the symbol scheme employed to label systems characterized with different rates and pressures. PC1 scores are significantly larger ($5\times$) compared to PC2, with PC1 capturing 98% of the variance in the entire dataset consisting of 49 systems. The largest variation in the pair orientation tensor components along PC1 is contributed by side atom pairs from high-pressure ($P = 636, 875, 955$ MPa) states; these pairs are not separable by rate or pressure. The end atom pairs from high-pressure states and the end-to-end atom pairs associated with low pressure ($P = 0.1, 100$ MPa) systems at low rates $\lesssim 10^8 \text{ s}^{-1}$ are among the pairs that contribute the next largest variation in $S_{\alpha\beta}^p$ components. These end-to-end atom pairs form clusters separable by rate and pressure. The backbone atom pairs, not separable by rate or pressure, are among the pairs with the next largest positive PC1 scores. A majority of atom pairs associated with high-pressure ($P \geq 400$ MPa) systems at strain rates $> 10^6 \text{ s}^{-1}$, including the end-to-end atom pairs, are grouped into a large mix of pairs that contribute in contrasting role with negative PC1 scores.

Fig. 10 **a** Behavior of atom pairs associated with a squalane molecule illustrated via dimension reduction of orientation tensors using PCA for all 49 sheared systems characterized with different pressures $P \in (0.1, 955)$ MPa and rates $\dot{\gamma} \in (10^6, 10^{10})$ s⁻¹. Each point represents one pair orientation tensor. Different symbols represent different systems identified with the pressure-rate (P - $\dot{\gamma}$) labels shown in the legend panel on the right. P - $\dot{\gamma}$ label with the same color mark the center of the data with respect to all the pairs associated with that system. Side, backbone, and end-to-end pairs are labeled with pair numbers in blue, olive green, and magenta colors, respectively. The

Examining the pressure and rate labels that define the center of the reduced $S_{\alpha\beta}^p$ data for each of the 49 systems shows the presence of 3 broad groups. Systems characterized with low rates and low pressures ($\dot{\gamma} \leq 2 \times 10^7$ s⁻¹ for 100 MPa and $\dot{\gamma} \leq 2 \times 10^8$ s⁻¹ for 0.1 MPa) form a clearly separate group with high and positive PC1 scores. The high-pressure systems ($P \geq 400$ MPa) at strain rates $> 10^6$

values in the brackets on the PC1 and PC2 axis labels are the variance ratios which signify the amount of variance captured along the respective axis. **b** Behavior of squalane under different P and $\dot{\gamma}$ conditions illustrated via the dimension reduction of all pairs and all orientation tensor components using PCA. The same color scheme noted in the legend on the right is used to represent all the 49 systems. **c** Rate dependence of the PC1 scores associated with systems at pressures P indicated in the legend in (c). PC1 scores are obtained from the scatter plot shown in (b) (Color figure online)

s⁻¹ form a distinct group with negative PC1 scores, indicating the contrasting nature of their contributions to variations in $S_{\alpha\beta}^p$ components. Finally, systems characterized with small positive PC1 scores between ~ 0 and ~ 0.15 (e.g., $\dot{\gamma} > 2 \times 10^8$ s⁻¹ for 0.1 MPa) form another group.

Our experiments with dimension reduction of six pair orientation tensor components to 2 dimensions for all 435

pairs reveal the important roles of the side atom pairs and the end-to-end atom pairs in assessing the correlation between changes in molecular alignment and shear thinning under increasing pressures. In addition to these studies, PCA can be used to process and reduce the entire $N = 2610$ dimensional data, consisting of 435 pairs and six orientation tensor components to 2 dimensions in order to reveal what features dominate the variations across all 2610 dimensions for all 49 systems. This dimension reduction results in 1 point for each system characterized with a given rate and pressure.

Figure 10b shows a single scatter plot of the reduction of the 2610-dimensional orientation tensor components associated with the shear flow of squalane for all strain rates and pressures (49 different systems). The symbol scheme is the same as noted in Fig. 10a. PC1 scores are 10× larger compared to PC2, with PC1 capturing 99% of the variance in the entire dataset consisting of all 49 systems. Figure 10b shares similar information as Fig. 10a but with more clarity. The aforementioned three groups of systems are more clearly identified in Fig. 10b. Low-pressure systems sheared at smaller strain rates are characterized with PC1 scores > 6 and clearly separate out as a group that can be associated with flow closer to the Newtonian regime. Within this group, it is easy to separate different systems. The high-pressure systems ($P \geq 400$ MPa) at strain rates $> 10^6$ s⁻¹ form a distinct group with negative PC1 scores between ~ 0 and ~ -3 . Most of the systems within this group are not easily separable as evident from the large mix of closely spaced pressure-rate labels. This group can be associated with non-Newtonian shear thinning dominated by a mechanism that does not rely on changing orientational order. Finally, systems characterized with small positive PC1 scores between ~ 0 and ~ 3 form another group. Within this group, the systems are relatively more separable and can be associated with non-Newtonian shear thinning dominated by changes in molecular alignment. The crossover from Newtonian regime to non-Newtonian shear thinning is clearly seen with increasing strain rate.

The dimension reduction tasks based on pair orientation tensors show that changes in molecular order in squalane are not correlated with strain rate for high-pressure systems even after considering all 435 pairs and all 6 $S_{\alpha\beta}^p$ components, as evident by the large mix of non-separable labels for these systems in Fig. 10b. This can also be observed by examining the PC1 scores for each rate and pressure obtained from Fig. 10b as a function of shear strain rate. Figure 10c shows that PC1 scores for low-pressure systems (0.1, 100 MPa) are strongly correlated with increasing rate. On the other hand, PC1 scores exhibit minimal variation with rate for the high-pressure systems. Figure 10c closely resembles the S_{yy} results shown in the order parameter plot (Fig. 3). This

similarity bolsters the conclusions drawn from the earlier study [9] using the orientation tensor components along flow and gradient direction associated with a single end-to-end atom pair. The utilization of all the six components of the orientation tensors for all 435 pairs reaffirms that the decrease in viscosity with rate for low pressures is strongly correlated with changes in the alignment of molecules. However, for high pressures, shear thinning occurs at saturated orientational order.

5 Discussion and Conclusion

The complexity of the high-dimensional output data originating from non-equilibrium molecular dynamics simulations can often make its analysis time consuming using conventional post-processing methods. We highlighted the capabilities of machine learning-based feature investigation tasks to expedite the analysis of this data and obtain reliable information. The approach was illustrated by examining the link between the rheological properties and the molecular-scale structure of squalane via the dimension reduction of the information encoded in all the six components of the orientation tensors of all 435 atom pairs associated with a squalane molecule. PCA was used to assess the correlation between strain rate and the evolution of the components of the pair orientation tensors $S_{\alpha\beta}^p$ for different pressures, and examine the contributions of distinct sets of atom pairs to the variation in the six components of the pair orientation tensors. t-SNE was used to more precisely assess the clustering of pair orientation tensor data by PCA and probe the crossover from Newtonian to non-Newtonian flow regimes.

Dimension reduction showed a clear, distinct grouping of atom pairs in the reduced dimension for systems characterized with small strain rates and pressures (0.1, 100 MPa), signaling a dramatic change in the flow behavior with increasing rate for a given pressure. This was related to the transition from Newtonian to non-Newtonian flow. The associated crossover rates predicted via dimension reduction can be used to determine the characteristic rate for the onset of shear thinning in Carreau and Eyring models.

Dimension reduction of 6-dimensional $S_{\alpha\beta}^p$ data in 2 dimensions showed that both the side atom pairs and the end-to-end atom pairs contribute significantly to variations in the orientation tensor components, but in contrasting roles. The end-to-end atom pairs dominated the largest variations in $S_{\alpha\beta}^p$ components for low-pressure systems (0.1, 100 MPa), and provided the clearest separation of the $S_{\alpha\beta}^p$ data with rate. A strong correlation between the shear rate and the PC1 scores was observed. On the other hand, the side atom pairs dominated the largest

variation in $S_{\alpha\beta}^p$ components for the high-pressure systems ($P \geq 400$ MPa). For these systems, no clear correlation between shear rate and PC1 scores was observed and the ability of the end-to-end atom pairs to separate data at different rates worsened with increasing pressure. The evolution in $S_{\alpha\beta}^p$ vs rate for most pairs associated with high-pressure systems saturated relatively quickly, implying a small and limited contribution to shear thinning due to changes in shear-driven molecular alignment.

Dimension reduction of the 2610-dimensional data consisting of all 435 pairs and six components of the orientation tensors for all strain rates and pressures reaffirmed the presence of 3 distinct groups of systems associated with the Newtonian flow, shear thinning with nearly saturated molecular order, and shear thinning dominated by changes in molecular alignment. Examining the PC1 scores for this dimension reduction task bolstered the conclusions drawn from the earlier study by Jadhao and Robbins [9] using the orientation tensor components along flow and gradient direction associated with a single end-to-end atom pair. Specifically, the utilization of all the six components of the orientation tensors for all 435 pairs showed that the decrease in viscosity with rate for low pressures is strongly correlated with changes in the alignment of molecules along the flow direction. However, for high pressures, shear thinning occurred at saturated orientational order.

Dimension reduction showed that the mechanisms responsible for shear thinning under high pressures, where squalane exhibits glassy flow, do not involve long-range molecular alignment such as that associated with the end-to-end atom pairs, but may involve the rearrangements of the side atom pairs. These rearrangements could be dominated by shear-assisted thermal activation [62, 63]. Other mechanisms such as those originating from the elasticity of the squalane molecules as represented in the Prandtl model may also explain the origins of the shear thinning behavior at high pressures [64]. New investigations utilizing different types of high-dimensional data are required to make progress toward a more precise diagnosis. For example, investigating the time evolution of the pair orientation tensors (averaged only over the number of molecules) can help preserve and quantify the rearrangements of different pairs with time. Further, enhancing the input information provided to the dimension reduction task by including the pressure and kinetic energy tensors can also assist in diagnosing the origins of the non-Newtonian shear flow under high pressures.

Dimension reduction methods explored in this work can be used to complement simulations aimed at describing the shear flow of other small-molecular liquids. In particular, studies of liquids such as nonadecane and 9-octylheptadecane, which have a similar number of carbon atoms as squalane, can help assess the correlation between shear thinning and molecular alignment for molecules of different architecture. These methods can also be extended to long-chain polymer systems to estimate how

much shear thinning may come from alignment [25, 56]. In many applications, EHL fluids contain a mixture of molecules of different architecture and length. There may be shear thinning correlated with the order of one species, say a long polymer, that is dissolved in a lower viscosity base oil. Application of dimension reduction methods to analyze the simulations of such fluid mixtures over a range of pressures and temperatures [65] could also be valuable. In general, we expect that the use of the dimension reduction methods to visualize high-dimensional information can complement simulations and experiments aimed at furnishing an improved understanding of the molecular-level mechanisms underlying the macroscopic rheological response of liquids.

Acknowledgements V.J. thanks Mark Robbins for mentoring him and for being a constant source of inspiration. We thank Geoffrey Fox for useful discussions.

Funding This research was conducted using computational resources at the Maryland Advanced Research Computing Center (MARCC) and the Big Red III supercomputing system at Indiana University.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Data Availability Our code and the experimental data used are publicly available on GitHub at <https://github.com/softmaterials/mflow/>

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Larson, R.G.: The Structure and Rheology of Complex Fluids, vol. 150. Oxford University Press, New York (1999)
2. Doi, M., Edwards, S.F.: The theory of polymer dynamics, vol. 73. Oxford University Press, Oxford (1988)
3. Adachi, J., Siebrits, E., Peirce, A., Desroches, J.: Computer simulation of hydraulic fractures. *Int. J. Rock Mech. Min. Sci.* **44**(5), 739–757 (2007)
4. Barbati, A.C., Desroches, J., Robisson, A., McKinley, G.H.: Complex fluids and hydraulic fracturing. *Annu. Rev. Chem. Biomol. Eng.* **7**, 415–453 (2016)
5. Spikes, H., Jie, Z.: History, origins and prediction of elastohydrodynamic friction. *Tribol. Lett.* **56**(1), 1–25 (2014)

6. Spikes, H., Zhang, J.: Reply to the comment by Scott Bair, Philippe Vergne, Punit Kumar, Gerhard Poll, Ivan Krupka, Martin Hartl, Wassim Habchi, Roland Larson on history, origins and prediction of elastohydrodynamic friction by spikes and JIE in tribology letters. *Tribol. Lett.* **58**(1), 17 (2015)
7. Ewen, J.P., Gattinoni, C., Zhang, J., Heyes, D.M., Spikes, H.A., Dini, D.: On the effect of confined fluid molecular structure on nonequilibrium phase behaviour and friction. *Phys. Chem. Chem. Phys.* **19**(27), 17883–17894 (2017)
8. Ewen, J.P., Heyes, D.M., Dini, D.: Advances in nonequilibrium molecular dynamics simulations of lubricants and additives. *Friction* **6**, 349–386 (2018)
9. Jadhao, V., Robbins, M.O.: Rheological properties of liquids under conditions of elastohydrodynamic lubrication. *Tribol. Lett.* **67**(3), 66 (2019)
10. Bair, S.: The high pressure rheology of some simple model hydrocarbons. *Proc. Inst. Mech. Eng. Part J J. Eng. Tribol.* **216**(3), 139–149 (2002)
11. Bair, S.: Reference liquids for quantitative elastohydrodynamics: selection and rheological characterization. *Tribol. Lett.* **22**(2), 197–206 (2006)
12. Bair, S.: *High-Pressure Rheology for Quantitative Elastohydrodynamics*. Elsevier Science, Amsterdam (2007)
13. Bair, S., Vergne, P., Kumar, P., Poll, G., Krupka, I., Hartl, M., Habchi, W., Larsson, R.: Comment on history, origins and prediction of elastohydrodynamic friction by spikes and JIE. *Tribol. Lett.* **58**(1), 16 (2015)
14. Eyring, H.: Viscosity, plasticity, and diffusion as examples of absolute reaction rates. *J. Chem. Phys.* **4**(4), 283–291 (1936)
15. Ewell, R.H., Eyring, H.: Theory of the viscosity of liquids as a function of temperature and pressure. *J. Chem. Phys.* **5**(9), 726–736 (1937)
16. Carreau, P.J.: Rheological equations from molecular network theories. *Trans. Soc. Rheol.* **16**(1), 99–127 (1972)
17. Evans, D.J., Morriss, G.P.: Nonlinear-response theory for steady planar Couette flow. *Phys. Rev. A* **30**, 1528–1530 (1984)
18. Evans, D.J., Morris, G.P.: *Statistical Mechanics of Nonequilibrium Liquids*. Academic Press, New York (1990)
19. Khare, R., de Pablo, J., Yethiraj, A.: Rheological, thermodynamic, and structural studies of linear and branched alkanes under shear. *J. Chem. Phys.* **107**(17), 6956–6964 (1997)
20. Moore, J.D., Cui, S.T., Cummings, P.T., Cochran, H.D.: Lubricant characterization by molecular simulation. *AIChE J.* **43**(12), 3260–3263 (1997)
21. Cui, S., Cummings, P., Cochran, H., Moore, J., Gupta, S.: Nonequilibrium molecular dynamics simulation of the rheology of linear and branched alkanes. *Int. J. Thermophys.* **19**(2), 449–459 (1998)
22. McCabe, C., Cui, S., Cummings, P.T., Gordon, P.A., Saeger, R.B.: Examining the rheology of 9-octylheptadecane to giga-pascal pressures. *J. Chem. Phys.* **114**(4), 1887–1891 (2001)
23. Kroger, M., Loose, W., Hess, S.: Rheology and structural changes of polymer melts via nonequilibrium molecular dynamics. *J. Rheol.* **37**(6), 1057–1079 (1993)
24. Jabbarzadeh, A., Atkinson, J., Tanner, R.: Effect of molecular shape on rheological properties in molecular dynamics simulation of star, H, comb, and linear polymer melts. *Macromolecules* **36**(13), 5020–5031 (2003)
25. Baig, C., Mavrantzas, V.G., Kroger, M.: Flow effects on melt structure and entanglement network of linear polymers: results from a nonequilibrium molecular dynamics simulation study of a polyethylene melt in steady shear. *Macromolecules* **43**(16), 6886–6902 (2010)
26. Khabaz, F., Khare, R.: Effect of chain architecture on the size, shape, and intrinsic viscosity of chains in polymer solutions: a molecular simulation study. *J. Chem. Phys.* **141**(21), 214904 (2014)
27. Kioupis, L.I., Maginn, E.J.: Molecular simulation of poly- α -olefin synthetic lubricants: impact of molecular architecture on performance properties. *J. Phys. Chem. B* **103**(49), 10781–10790 (1999)
28. Moore, J.D., Cui, S.T., Cochran, H.D., Cummings, P.T.: Rheology of lubricant basestocks: a molecular dynamics study of C30 isomers. *J. Chem. Phys.* **113**(19), 8833–8840 (2000)
29. McCabe, C., Cui, S., Cummings, P.T.: Characterizing the viscosity-temperature dependence of lubricants by molecular simulation. *Fluid Phase Equilib.* **183**, 363–370 (2001)
30. Bair, S., McCabe, C., Cummings, P.T.: Comparison of nonequilibrium molecular dynamics with experimental measurements in the nonlinear shear-thinning regime. *Phys. Rev. Lett.* **88**, 058302 (2002)
31. Jadhao, V., Robbins, M.O.: Probing large viscosities in glass-formers with nonequilibrium simulations. *Proc. Natl. Acad. Sci. U.S.A.* **114**(30), 7952–7957 (2017)
32. Galvani, M., Robbins, M.O.: Determination of pressure-viscosity relation of 2,2,4-trimethylhexane by all-atom molecular dynamics simulations. *Fluid Phase Equilib.* **495**, 28–32 (2019)
33. O'Connor, T.C., Andzelm, J., Robbins, M.O.: Airebo-m: a reactive model for hydrocarbons at extreme pressures. *J. Chem. Phys.* **142**(2), 024903 (2015)
34. Lin, L., Kedzierski, M.A.: Density and viscosity of a polyol ester lubricant: measurement and molecular dynamics simulation. *Int. J. Refrig.* **118**, 188–201 (2020)
35. Jadhao, V., Robbins, M.O.: Reply to bair: Crossover to arrhenius behavior at high viscosities in squalane. *Proc. Natl. Acad. Sci. U.S.A.* **114**(42), E8807–E8808 (2017)
36. Ferguson, A.L.: Machine learning and data science in soft materials engineering. *J. Phys. Condens. Matter* **30**(4), 043002 (2017)
37. Schoenholz, S.S.: Combining machine learning and physics to understand glassy systems. *J. Phys. Conf. Ser.* **1036**(1), 012021 (2018)
38. Haghighatdari, M., Hachmann, J.: Advances of machine learning in molecular modeling and simulation. *Curr. Opin. Chem. Eng.* **23**, 51–57 (2019)
39. Wang, J., Gayatri, M.A., Ferguson, A.L.: Mesoscale simulation and machine learning of asphaltene aggregation phase behavior and molecular assembly landscapes. *J. Phys. Chem. B* **121**(18), 4923–4944 (2017). <https://pubs.acs.org/doi/abs/10.1021/acs.jpcc.7b02574>
40. Wang, J., Olsson, S., Wehmeyer, C., Pérez, A., Charron, N., De Fabritiis, G., Noé, F., Clementi, C.: Machine learning of coarse-grained molecular dynamics force fields. *ACS Central Science* **5**(5), 755–767 (2019)
41. Kadupitiya, J., Fox, G.C., Jadhao, V.: Simulating molecular dynamics with large timesteps using recurrent neural networks. *arXiv preprint* (2020). [arXiv:2004.06493](https://arxiv.org/abs/2004.06493)
42. Sun, Y., DeJaco, R.F., Siepmann, J.I.: Deep neural network learning of complex binary sorption equilibria from molecular simulation data. *Chem. Sci.* **10**(16), 4377–4388 (2019)
43. Moradzadeh, A., Aluru, N.R.: Molecular dynamics properties without the full trajectory: a denoising autoencoder network for properties of simple liquids. *J. Phys. Chem. Lett.* **10**(24), 7568–7576 (2019)
44. Kadupitiya, J., Sun, F., Fox, G., Jadhao, V.: Machine learning surrogates for molecular dynamics simulations of soft materials. *J. Comput. Sci.* **42**, 101107 (2020)
45. Kadupitiya, J., Fox, G.C., Jadhao, V.: Machine learning for performance enhancement of molecular dynamics simulations. In: *International Conference on Computational Science*, pp. 116–130 (2019). https://link.springer.com/chapter/10.1007/978-3-030-22741-8_9

46. Kadupitiya, J., Fox, G.C., Jadhao, V.: Machine learning for parameter auto-tuning in molecular dynamics simulations: Efficient dynamics of ions near polarizable nanoparticles. *Int. J. High Perform. Comput. Appl.* (2020). <https://doi.org/10.1177/1094342019899457>
47. Jadhao, V., Kadupitiya, J.: Integrating machine learning with hpc-driven simulations for enhanced student learning. In: 2020 IEEE/ACM Workshop on Education for High-Performance Computing (EduHPC), pp. 25–34. IEEE (2020)
48. Fox, G., Glazier, J.A., Kadupitiya, J., Jadhao, V., et al.: Learning everywhere: Pervasive machine learning for effective high-performance computation. In: IEEE IPDPS Workshops, pp. 422–429 (2019). <https://doi.org/10.1109/IPDPSW.2019.00081>
49. Bejagam, K.K., Singh, S.K., Ahn, R., Deshmukh, S.A.: Unraveling the conformations of backbone and side chains in thermosensitive bottlebrush polymers. *Macromolecules* **52**(23), 9398–9408 (2019)
50. Schoenholz, S.S., Cubuk, E.D., Kaxiras, E., Liu, A.J.: Relationship between local structure and relaxation in out-of-equilibrium glassy systems. *Proc. Natl Acad. Sci. U.S.A.* **114**(2), 263–267 (2017)
51. Mahmoudabadbozchelou, M., Caggioni, M., Shahsavari, S., Hartt, W.H., Em Karniadakis, G., Jamali, S.: Data-driven physics-informed constitutive metamodeling of complex fluids: a multifidelity neural network (MFNN) framework. *J. Rheol.* **65**(2), 179–198 (2021)
52. Mondello, M., Grest, G.S.: Molecular dynamics of linear and branched alkanes. *J. Chem. Phys.* **103**(16), 7156–7165 (1995)
53. Mondello, M., Grest, G.S., Garcia, A.R., Silbernagel, B.G.: Molecular dynamics of linear and branched alkanes: simulations and nuclear magnetic resonance results. *J. Chem. Phys.* **105**(12), 5208–5215 (1996)
54. Plimpton, S.: Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**(1), 1–19 (1995)
55. Kroger, M., Hess, S.: Viscoelasticity of polymeric melts and concentrated solutions. the effect of flow-induced alignment of chain ends. *Physica A Stat. Mech. Appl.* **195**(3), 336–353 (1993)
56. Barsky, S., Robbins, M.O.: Bulk and interfacial shear thinning of immiscible polymers. *Phys. Rev. E* **65**, 021808 (2002)
57. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. *Chemom. Intell. Lab. Syst.* **2**(1–3), 37–52 (1987)
58. Abdi, H., Williams, L.J.: Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2**(4), 433–459 (2010)
59. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
60. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. *J. Mach. Learn. Res.* **9**(11), (2008)
61. Van Der Maaten, L., Postma, E., Van den Herik, J.: Dimensionality reduction: a comparative. *J. Mach. Learn. Res.* **10**(66–71), 13 (2009)
62. Spikes, H.: Stress-augmented thermal activation: tribology feels the force. *Friction* **6**(1), 1–31 (2018)
63. Baker, D.L., Reynolds, M., Masurel, R., Olmsted, P.D., Mattsson, J.: Chain-length, flexibility and the glass transition of polymers. *arXiv preprint* (2019). [arXiv:1911.13278](https://arxiv.org/abs/1911.13278)
64. Müser, M.H.: Shear thinning in the Prandtl model and its relation to generalized Newtonian fluids. *Lubricants* **8**(4), 38 (2020)
65. Punchard, B.J., Adolf, D.: Pressure and temperature dependence of the dilute solution segmental dynamics of anthracene-labeled polyisoprene. *Macromolecules* **35**(8), 3281–3287 (2002)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.