



Keywords: small-angle scattering; machine learning; Gaussian process regressors; Monte Carlo methods; polymers.

Machine learning inversion from scattering for mechanically driven polymers

Lijie Ding,^a Chi-Huan Tung,^a Bobby G. Sumpter,^b Wei-Ren Chen^{a*} and Changwoo Do^{a*}

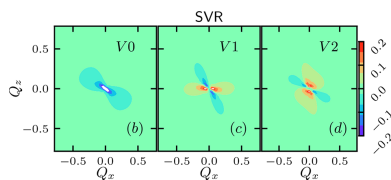
^aNeutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA, and ^bCenter for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. *Correspondence e-mail: chenw@ornl.gov, doc1@ornl.gov

A machine learning inversion method is developed for analyzing scattering functions of mechanically driven polymers and extracting the corresponding feature parameters, which include energy parameters and conformation variables. The polymer is modeled as a chain of fixed-length bonds constrained by bending energy, and it is subject to external forces such as stretching and shear. We generate a data set consisting of random combinations of energy parameters, including bending modulus, stretching and shear force, along with Monte Carlo-calculated scattering functions and conformation variables such as end-to-end distance, radius of gyration and off-diagonal component of the gyration tensor. The effects of the energy parameters on the polymer are captured by the scattering function, and principal component analysis ensures the feasibility of the machine learning inversion. Finally, we train a Gaussian process regressor using part of the data set as a training set and validate the trained regressor for inversion using the rest of the data. The regressor successfully extracts the feature parameters.

1. Introduction

Machine learning (ML) (Murphy, 2012; Carleo *et al.*, 2019) has emerged as a powerful tool for data analysis, enabling the extraction of patterns, trends and insights from large complex data sets. Its ability to automate the discovery of meaningful relationships within data has helped to advance numerous fields, including scattering analysis (Chang *et al.*, 2022). ML techniques can be used for the rapid interpretation of underlying material properties and structural parameters according to complex scattering data. This technique has been applied to various systems including colloids (Chang *et al.*, 2022; Huang *et al.*, 2023; Tung *et al.*, 2023), polymers (Tung *et al.*, 2022; Ding *et al.*, 2025a) and lyotropic lamellar systems (Tung *et al.*, 2024b).

Polymers are ubiquitous in nature and play a pivotal role in everyday life and in numerous industry settings (De Gennes, 1979; De Gennes, 1990; Sperling, 2015). Understanding the physics of polymers can help us to better design and engineer new materials for different applications. The polymers' response to external forces is often of interest as the mechanical properties of the polymer can thus be revealed (Wang & Wang, 2011; Li & Wang, 2016; Smith *et al.*, 1999; Schroeder *et al.*, 2005). Due to the small physical size of most polymers, scattering experiments (Murphy *et al.*, 2020), such as X-ray (Chu & Hsiao, 2001) or neutron (Shibayama, 2011; Chen, 1986) scattering, are commonly employed to probe their structure and dynamics at the molecular level. The scattering function measured by these experiments provides indirect but



OPEN ACCESS

Published under a CC BY 4.0 licence

valuable information about the polymer's conformation and behavior under mechanical stress. Recent advancements in RheoSANS (rheological small-angle neutron scattering; Murphy *et al.*, 2020) and sample environments have enabled the application of external forces that are comparable to the bending energy of the polymer using flow cells. A Monte Carlo (MC) (Krauth, 2006) method that we recently developed (Ding *et al.*, 2024) has enabled the theoretical study of mechanically driven polymers, which are semiflexible polymers under external forces including stretching and shear, and the calculation of scattering functions compatible with scattering experiments. Unlike the one-dimensional scattering profile found in isotropic systems, the scattering function of a mechanically driven polymer is not isotropic due to the existence of non-isotropic forces (Huang *et al.*, 2025), rendering the existing models including Gaussian chain (Debye, 1947), worm-like chain (Kholodenko, 1993) and flexible cylinder (Pedersen & Schurtenberger, 1996) unsuitable for scattering analysis.

The lack of a scattering analysis technique prevents us from extracting physical parameters at the molecular level from mechanically driven polymers using small-angle scattering experiments (Huang *et al.*, 2025). For this we turn instead to ML for a practical solution. ML techniques like Gaussian process regression (GPR) (Williams & Rasmussen, 2006) have been used to extract physical parameters directly from scattering data (Chang *et al.*, 2022). Generative models utilizing a variational autoencoder (Tung *et al.*, 2024a; Ding *et al.*, 2025b) have been combined with a regression model to fit the scattering function. The GPR method is more straightforward yet has only been applied to 1D scattering data so far, while the scattering data for a mechanically driven polymer are naturally 2D due to the anisotropy introduced by the external forces. In this paper, we extend the GPR to 2D scattering and apply the ML inversion technique (Chang *et al.*, 2022) to map the scattering function onto inversion targets or feature parameters of the mechanically driven polymers. We generate a data set for training and testing using the MC simulation that we previously developed. The effects of energy parameters such as bending, stretching and shear on the scattering function are well reflected independently and the corresponding polymer deformation is well captured by the calculated scattering function. The feasibility of the proposed ML inversion framework is validated by principal component analysis, which also provides the characteristic orientation of the scattering function as a byproduct. Excellent agreement between the ML-extracted feature parameters and the MC reference values is achieved, showing good accuracy for our approach.

2. Method

We model the polymer as a chain of N connected bonds with fixed length l_b . The tangent of bond i is $\mathbf{t}_i \equiv (\mathbf{r}_{i+1} - \mathbf{r}_i)/l_b$, where $\mathbf{r}_i$ is the position of the joint connecting bonds $i - 1$ and i . We fix one end of the polymer at the origin. The polymer energy is given by

$$E = \sum_{i=0}^{N-2} \frac{\kappa}{2} \frac{(\mathbf{t}_{i+1} - \mathbf{t}_i)^2}{l_b} - \sum_{i=0}^{N-1} (\gamma z_i + f) (l_b \mathbf{t}_i \cdot \mathbf{x}), \quad (1)$$

where κ is the bending modulus, f is the stretching force applied in the x direction, γ is the shear ratio along the z direction, $z_i = \mathbf{r}_i \cdot \mathbf{z}$ is the z component of the position of joint i and $(\mathbf{t}_i \cdot \mathbf{x})$ is the x component of the bond tangent $\mathbf{t}_i$. A hard-sphere interaction between polymer joints, with a sphere radius $l_b/2$, was used to account for self-avoidance of the polymer.

We sample the configuration space of the polymer using MC and calculate the scattering function and conformation variables of the polymer. We then use GPR to achieve a mapping from the scattering function to the system parameters and conformation variables.

2.1. Monte Carlo simulation

We simulate the polymer under different system parameters using the Markov chain MC method (Ding *et al.*, 2024) that we previously developed, where each configuration of the polymer is generated by updating the previous one. Two types of non-local moves are used for updating the polymer: crankshaft and pivot. Crankshaft executes a random rotation of an inner sub-chain of the polymer, while pivot rotates the sub-chain including the end. Details of the simulations can be found in our previous work (Ding *et al.*, 2024). From the simulations, the scattering function and other conformation variables, including end-to-end distance, radius of gyration and off-diagonal component of the gyration tensor, were computed. The scattering function is defined as (Chen, 1986)

$$I(\mathbf{Q}) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \exp[-i\mathbf{Q} \cdot (\mathbf{r}_i - \mathbf{r}_j)], \quad (2)$$

where $\mathbf{Q}$ is the scattering vector [$Q = (4\pi/\lambda) \sin(\theta/2)$, where θ is the scattering angle and λ is the wavelength of the incident radiation] and N is the total number of segments. In practice, a projection of $I(\mathbf{Q})$ onto a specific plane is collected in scattering experiments. Since the force field is applied in the (x, z) plane or flow-velocity gradient plane, we calculate the two-dimensional $I_{xz}(\mathbf{Q}) = I(Q_x, Q_y = 0, Q_z)$ accordingly. In addition, the end-to-end distance is defined as $R^2 = |\mathbf{r}_0 - \mathbf{r}_{N-1}|^2$, the radius of gyration is $R_g^2 = \frac{1}{2} \langle |\mathbf{r}_i - \mathbf{r}_j|^2 \rangle_{i,j}$ and the xz component of the gyration tensor is $R_{xz} = \frac{1}{2} \langle (x_i - z_j)^2 \rangle_{i,j}$, with $\langle \dots \rangle_{i,j}$ denoting the average over all pairs of joints.

2.2. Gaussian process regression

To obtain a mapping from the scattering function $\mathbf{x} = I_{xz}(\mathbf{Q})$ to inversion targets $\mathbf{y}$ including both system parameters and conformation variables, we train a GPR (Williams & Rasmussen, 2006) by feeding training data $\{I_{xz}^{\text{train}}(\mathbf{Q})\}$ containing scattering functions calculated with various system parameters (κ, f, γ). We define the prior on the regression function as a Gaussian process $g(\mathbf{x}) \simeq GP[m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')]]$, where $m(\mathbf{x})$ is the prior mean function and $k(\mathbf{x}, \mathbf{x}')$ is the covariance function or kernel. Given a test data set

$\mathbf{X}_* = \{I_{xz}^{\text{test}}(\mathbf{Q})\}$, the goal of the regressor is to estimate $\mathbf{Y}_* = g(\mathbf{X}_*)$. The joint distribution for a Gaussian process is

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{Y}_* \end{pmatrix} \simeq \mathcal{N}\left(\begin{bmatrix} m(\mathbf{X}) \\ m(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} k(\mathbf{X}, \mathbf{X}) & k(\mathbf{X}, \mathbf{X}_*) \\ k(\mathbf{X}_*, \mathbf{X}) & k(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix}\right), \quad (3)$$

where we use a constant prior mean $m(\mathbf{x})$ and a linear combination of a radial basis function (Gaussian) kernel and white noise for the kernel $k(\mathbf{x}, \mathbf{x}') = \exp[-d(\mathbf{x}, \mathbf{x}')^2/(2l)] + \sigma\delta(\mathbf{x}, \mathbf{x}')$. Here, $d(\mathbf{x}, \mathbf{x}') = |\mathbf{x} - \mathbf{x}'|$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}'$, l is the correlation length, σ is the variance of observational noise, and δ is the Kronecker delta function. l and σ are the hyperparameters for the regression and can be obtained by training.

3. Results

We prepare the training $\{I_{\text{train}}(\mathbf{Q})\}$ and test $\{I_{\text{test}}(\mathbf{Q})\}$ sets by carrying out MC simulations of the polymer chains with various combinations of energy parameters: bending modulus κ , stretching force f and shear rate γ . The scattering function and conformation variables were measured for each simulation. We use natural units in our simulation such that lengths are measured in units of the bond length l_b and energies are measured in units of thermal noise $k_B T$. Prior to training, we first study the effect of energy parameters on the scattering function and then validate the feasibility of inversion using principal component analysis. Finally, we train our GPR and compare the ML-calculated inversion targets with values calculated using MC.

3.1. Scattering function of the polymers

In order for the GPR to achieve mapping from the scattering function to the inversion targets, the scattering function must reflect the changes in the inversion targets, *i.e.* the energy parameters. These results are demonstrated in Figs. 1 and 3, where the scattering function at various bending modulus κ , stretching force f and shear rate γ are shown.

The bending modulus κ determines the persistence length of the polymer. A longer persistence length makes the polymer more rod-like, thus lowering the scattering intensity $I_{xz}(\mathbf{Q})$ at larger $Q = |\mathbf{Q}|$. Fig. 1 shows $I_{xz}(\mathbf{Q})$ at different κ ; the contour of $I_{xz}(\mathbf{Q})$ shows circular symmetry, indicating isotropy of the polymer system in the absence of external forces. The

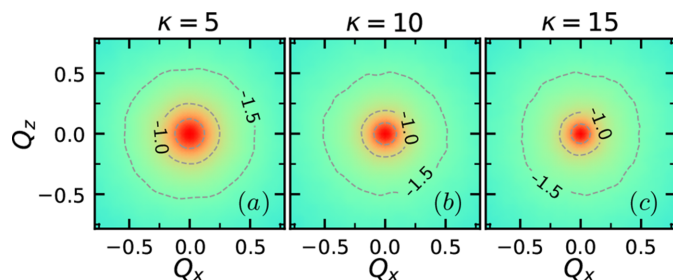


Figure 1
 $I_{xz}(\mathbf{Q})$ of a semiflexible chain with $L = 200$ in its quiescent state with bending modulus $\kappa = 5, 10$ and 15 .

ring of contour level also shrinks as the bending modulus κ increases from Fig. 1(a) to Fig. 1(c), consistent with our intuition about the effect of κ on the persistence length.

When external forces are applied, the polymer deforms accordingly. Fig. 2 shows sample configurations of the polymer under different stretching f and shear γ . When stretching f is only applied along the x direction, the polymer extends along the x direction as shown in Figs. 2(d) and 2(g). Figs. 2(b) and 2(c) show that the polymer extends towards the xz direction in a convex manner when only the shear γ is applied. Combining the stretching force and shear rate, the polymer behaves like something in the middle, such that an increasing stretching force f pulls the polymer more towards the x direction [compare Figs. 2(b), 2(e) and 2(f)]. These deformations are also reflected in the scattering function. The anisotropic behavior of a polymer should deform the circular symmetric shape of $I_{xz}(\mathbf{Q})$.

Fig. 3 shows the scattering function $I_{xz}(\mathbf{Q})$ for the polymers corresponding to Fig. 2. The contour of the scattering function evolves into an oval and then a dumbbell shape as the applied force increases. $I_{xz}(\mathbf{Q})$ at high Q decreases with the increasing magnitude of stretching f and shear γ , reflecting an increase in the radius of gyration due to straightening. On the other hand, the ratio between f and γ affects the orientation of the $I_{xz}(\mathbf{Q})$ contour. For pure stretching, the contour of $I_{xz}(\mathbf{Q})$ extends along the z direction, indicating elongation of the polymer along the x direction. In contrast, pure shear makes the contour of $I_{xz}(\mathbf{Q})$ extend along the $-xz$ direction, reflecting

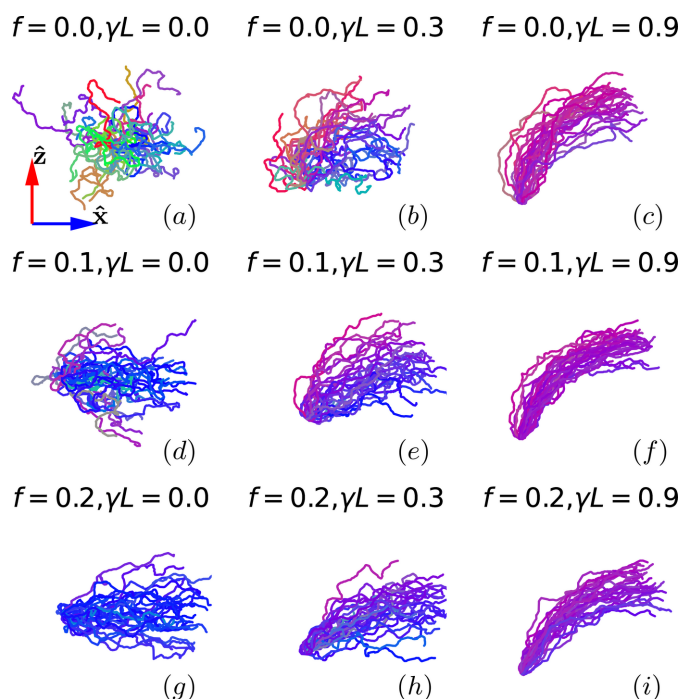
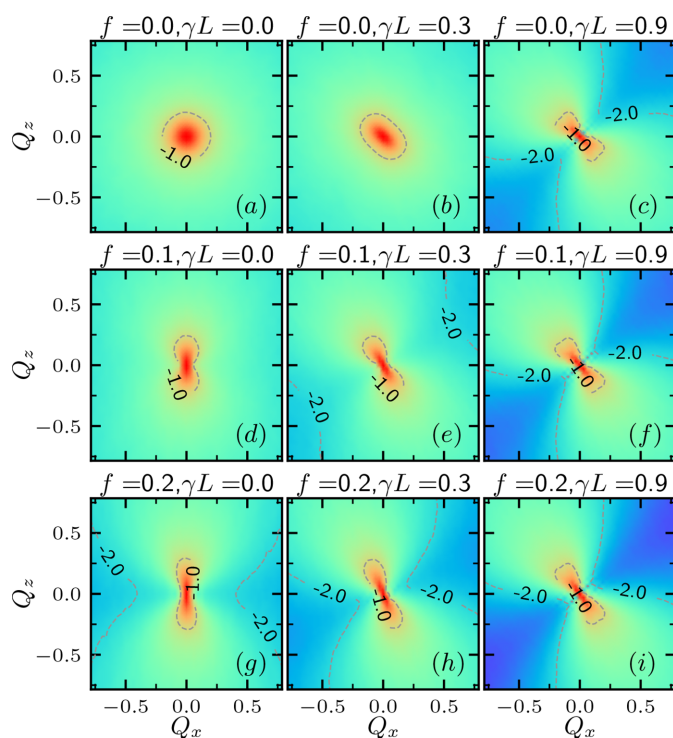


Figure 2
Sample configurations of a semiflexible chain with $L = 200$ and $\kappa = 10$ with various combinations of stretching and shear $(f, \gamma) = (0, 0.1, 0.2) \times (0, 0.3, 0.9)$. The color corresponds to the end-to-end orientation in the xz plane. The system is symmetric about $\pm xz$ for panels (b) and (c), where $f = 0$, $\gamma \neq 0$; these configurations are flipped to the xz direction for better visualization.


Figure 3

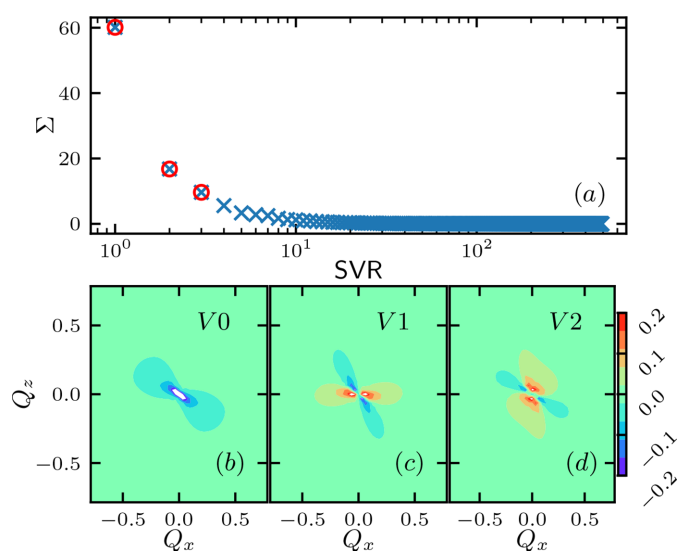
Scattering function $I_{xz}(\mathbf{Q})$ of a semiflexible chain with $L = 200$ and $\kappa = 10$ with various combinations of stretching and shear $(f, \gamma) = (0, 0.1, 0.2) \times (0, 0.3, 0.9)$.

the elongation of the polymer along the xz direction. Applying and increasing the shear rate on a polymer under stretching, as shown in Figs. 3(g), 3(h) and 3(i), rotates the orientation of the dumbbell-shaped contour towards the $-xz$ direction.

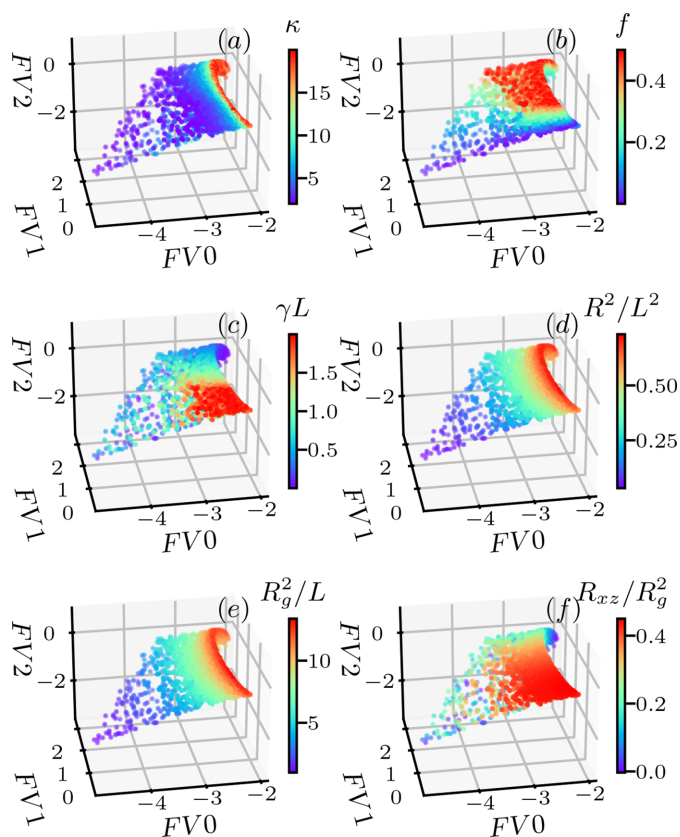
3.2. Feasibility of ML inversion

Due to the significant difference of the effect on scattering functions induced by different energy parameters, we anticipate that the difference in energy parameters can be distinguished from the scattering function using GPR. To validate the feasibility of such inversion numerically, we generate 1680 random combinations of (κ, f, γ) , in which $\kappa \simeq U(2, 20)$, $f \simeq U(0, 0.5)$, $\gamma L \simeq U(0, 2)$, where $U(a, b)$ is the uniform distribution in the interval $[a, b]$. We then run MC simulations to calculate the scattering function $I_{xz}(\mathbf{Q})$ of the polymer system at these energy parameters. Each $I_{xz}(\mathbf{Q})$ is calculated for 2601 = 51×51 different (Q_x, Q_z) , where $Q_x, Q_z \in [-50\pi/L, 50\pi/L]$, uniformly placed on the 51×51 grid. These $I_{xz}(\mathbf{Q})$ are then flattened to 2601-dimensional vectors and arranged into a 1680×2601 matrix $\mathbf{F}$. Following a similar ML inversion framework (Chang *et al.*, 2022), $\mathbf{F}$ is then decomposed into $\mathbf{F} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ using singular value decomposition (Strang, 2022), such that $\mathbf{U}$ is 1680×1680 , $\mathbf{\Sigma}$ is 1680×2601 and $\mathbf{V}$ is 2601×2601 . The singular value matrix $\mathbf{\Sigma}^2$ is diagonal and its entries are proportional to the variance of the data set $\mathbf{F}$ projected onto the corresponding principal axis (Zhu & Ghodsi, 2006), which is given by the singular vector $\mathbf{V}$.

Fig. 4(a) shows the diagonal entry value of $\mathbf{\Sigma}$ versus its singular value rank (SVR). As the SVR increases, its corresponding


Figure 4

Singular value decomposition of the scattering function data set. (a) Singular value Σ versus singular value rank. Values with top-three rank are highlighted with red circles. (b)–(d) First three singular vectors.


Figure 5

Distribution of various inversion features of the training data in the singular value space. (a) Bending modulus κ . (b) Stretching force f . (c) Contour length normalized shear γL . (d) End-to-end distance scaled by contour length squared R^2/L^2 . (e) Radius of gyration squared scaled by contour length R_g^2/L . (f) Off-diagonal xz component of gyration tensor R_{xz}/R_g^2 .

value decreases rapidly, indicating that the variations in $I_{xz}(\mathbf{Q})$ are dominated by the first few singular vectors of lower rank. Figs. 4(b)–4(d) show the first three single vectors, which give a characteristic basis for $I_{xz}(\mathbf{Q})$.

By projecting the data set $\mathbf{F}$ onto the first three singular vectors V_0 , V_1 and V_2 , the (FV_0, FV_1, FV_2) coordinates provide a good proxy of $\mathbf{F} = \{I_{xz}(\mathbf{Q})\}$. Fig. 5 shows the distribution of the six inversion targets in the (FV_0, FV_1, FV_2) space. Three of these are the energy parameters – bending modulus κ , stretching force f and shear rate γ – and the other three are conformation variables – end-to-end distance R^2 , radius of gyration R_g^2 and off-diagonal xz component of the gyration tensor R_{xz} . In this (FV_0, FV_1, FV_2) space, each point corresponds to one $I_{xz}(\mathbf{Q})$ in $\mathbf{F}$, and the color represents the corresponding value of the inversion target. From the color distribution, we note that the inversion targets, feature variables, are all well spread out in (FV_0, FV_1, FV_2) space, indicating that a smooth and continuous mapping between $I_{xz}(\mathbf{Q})$ and the inversion target can be obtained, thus validating the feasibility of the ML inversion.

3.3. ML inversion of feature variables

To illustrate the inversion of feature parameters $(\kappa, f, \gamma, R^2, R_g^2, R_{xz})$ from scattering functions $I_{xz}(\mathbf{Q})$, we

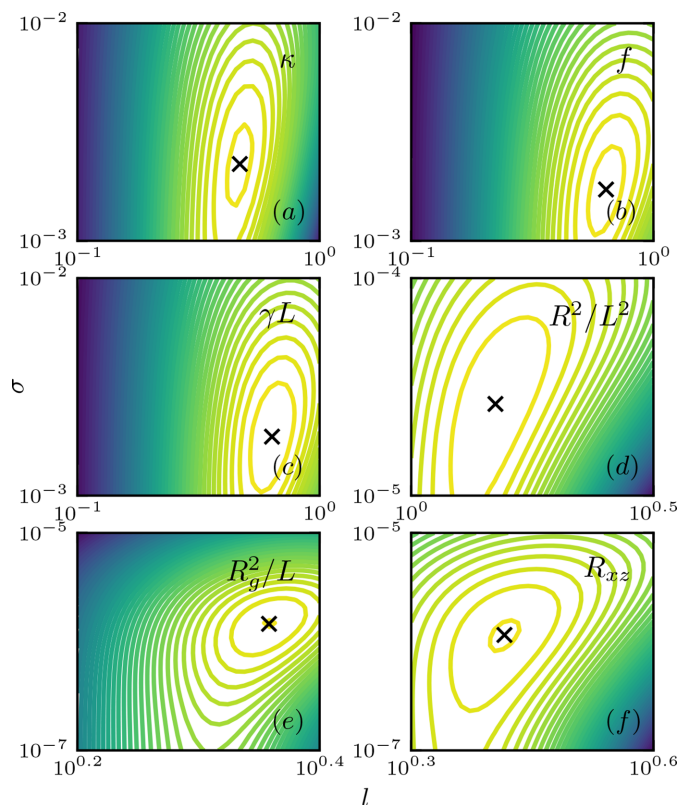


Figure 6

Log marginal likelihood surface of hyperparameters l and σ for various inversion features, with optimized values marked with black crosses. (a) Bending modulus κ . (b) Stretching force f . (c) Contour length normalized shear γL . (d) End-to-end distance scaled by contour length squared R^2/L^2 . (e) Radius of gyration squared scaled by contour length R_g^2/L . (f) Off-diagonal xz component of gyration tensor R_{xz}/R_g^2 .

partition the total data set $\mathbf{F} = \{I_{xz}(\mathbf{Q})\}$ into two sets: a training set $\{I_{xz}^{\text{train}}(\mathbf{Q})\}$ consisting of 70% of $\mathbf{F}$, and a test set $\{I_{xz}^{\text{test}}(\mathbf{Q})\}$ consisting of the other 30% of $\mathbf{F}$. We use the training set to obtain the optimized hyperparameters (l, σ) through gradient descent on the log marginal likelihood landscape. We do this for the kernel for each feature parameter individually, and then use the trained GPR with the optimized (l, σ) to predict the feature parameters of the test set from their $I_{xz}(\mathbf{Q})$. The *scikit-learn* Gaussian process library (Pedregosa *et al.*, 2011) was used for the training and inversion.

The log marginal likelihood of the prior is used as the cost function for optimizing the hyperparameters (l, σ) (Williams & Rasmussen, 2006). Fig. 6 shows the log marginal likelihood contour in (l, σ) space for each feature parameter or inversion target. The optimized (l, σ) are obtained by gradient descent and shown in Table 1. The contours in Fig. 6 show unimodal convex patterns, which further suggest the reliability of the trained hyperparameters. While the optimized hyperparameters (l, σ) differ for each inversion target, two scales of correlation length and noise level emerge. The optimized l and σ for all the energy parameters have the same order of magnitude, which is also true for the conformation parameters but with a higher order of magnitude, indicating that the scattering function is more sensitive to the variation in energy parameters compared with conformation change.

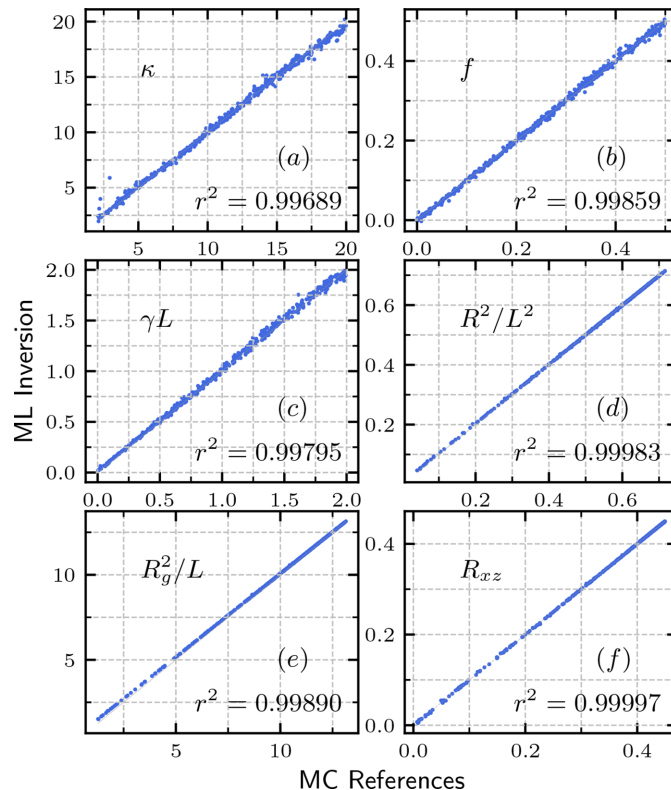


Figure 7

Comparison between the feature variables extracted from the scattering function $I(\mathbf{Q})$ using GPR and their corresponding computational references calculated using MC simulations. Coefficient of determination, r^2 scores, are indicated at the bottom of each panel.

Table 1
Optimized hyperparameters for each feature, obtained from maximum log marginal likelihood.

	l	σ
κ	4.6828×10^{-1}	2.2548×10^{-3}
f	6.3714×10^{-1}	1.7219×10^{-3}
γL	6.3591×10^{-1}	1.8671×10^{-3}
R^2/L^2	1.4921	2.6388×10^{-5}
R_g^2/L	2.2814	1.4582×10^{-6}
R_{xz}	2.6027	1.1527×10^{-6}

Finally, we use the scattering function from the test set $\{I_{xz}^{\text{test}}(\mathbf{Q})\}$ as input to the trained GPR and calculate the feature parameters ($\kappa, f, \gamma, R^2, R_g^2, R_{xz}$) as ML inversion from $I_{xz}(\mathbf{Q})$. Fig. 7 shows a comparison between the GPR-predicted feature parameters and the MC references. All of the data lie close to the diagonal line, with an r^2 score, coefficient of determination, close to 1. The high precision of the inversion shows the power of our ML approach for extracting important system information from the scattering function.

4. Summary

In summary, we have applied an ML inversion method to extract feature parameters from the scattering data of mechanically driven polymers. The ML inversion framework was trained on a theoretically calculated data set of a polymer system that is determined by several energy parameters: bending modulus κ , stretching force f and shear rate γ . The inversion targets included these energy parameters and also conformation variables such as the end-to-end distance R^2 , the radius of gyration R_g^2 and the off-diagonal component of the gyration tensor R_{xz} . The scattering function $I_{xz}(\mathbf{Q})$ of the polymer under different energy parameters was calculated using an MC method that we developed previously (Ding *et al.*, 2024). We have demonstrated the feasibility of the ML inversion by carrying out principal component analysis of the data set $\mathbf{F} = \{I_{xz}(\mathbf{Q})\}$ and investigated the distribution of feature parameters by projecting the data set $\mathbf{F}$ to a three-dimensional singular vector space. The GPR was trained and validated, showing that inversion of the feature parameters can be achieved with high precision.

The versatility of our method promotes its application to the inversion analysis of polymer systems characterized by different intrinsic interactions or under other external forces. For instance, the polymer chains are often charged, in which case instead of using a single parameter, the bending modulus, the interaction between monomers on the polymer can be modeled by the two-parameter Yukawa interaction (Robbins *et al.*, 1988). The sample environment of the RheoSANS experiments can introduce nonuniform shear flow like Hagen–Poiseuille flow (Batchelor, 2000). More complicated polymer systems, including polymer brushes (Feng & Huang, 2018), star polymers (Ren *et al.*, 2016) and polymer melts (Kremer & Grest, 1990), are also of interest. Modification of the MC simulation can be made accordingly and ML inversion analysis similar to this work can be carried out.

While our approach is general, inversion from scattering depends on the specific physical model. Just like traditional analysis with model functions relies on how rigorously the model functions were derived and are capable of representing the actual scattering data, our ML approach also relies on the accuracy of the simulated training data for representing the actual experimental system. If the experimental system differs significantly in these physical characteristics from our simulation, the inversion model would need to be retrained on more appropriate simulation data.

We note that the inversion method requires the input scattering function to have the same $\mathbf{Q}$ grid as the training set, which can lead to interpolation of the experimental data in practice. Recent developments in ML (Liu *et al.*, 2024; Tung *et al.*, 2025) show the feasibility of mapping from vectors to functions, which opens the possibility of an alternative method of scattering analysis. Instead of training the mapping from scattering data in discrete $\mathbf{Q}$ to feature parameters as an inversion, the new framework can learn the mapping from energy parameters to the scattering function in continuous $\mathbf{Q}$ values, which enables calculation of the scattering function that can then be used for a quick gradient descent optimization of the energy parameter directly. This approach can also be used to cross validate our inversion method. Finally, more detailed analysis of the performance of our method can be carried out, for example studying the behavior of the model on different sized grids of the scattering image and with different splits of the training and testing data.

Acknowledgements

We thank Jan-Michael Carrillo for fruitful discussions. This research was performed at the Spallation Neutron Source and the Center for Nanophase Materials Sciences, which are DOE Office of Science User Facilities operated by Oak Ridge National Laboratory.

Funding information

This research was sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the US Department of Energy. The ML aspects were supported by the US Department of Energy Office of Science, Office of Basic Energy Sciences, Data, Artificial Intelligence and Machine Learning at DOE Scientific User Facilities Program under award No. 34532. Monte Carlo simulations and computations used resources of the Oak Ridge Leadership Computing Facility, which is supported by the DOE Office of Science under contract DE-AC05-00OR22725.

References

Batchelor, G. K. (2000). *An introduction to fluid dynamics*. Cambridge University Press.

- Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L. & Zdeborová, L. (2019). *Rev. Mod. Phys.* **91**, 045002.
- Chang, M.-C., Tung, C.-H., Chang, S.-Y., Carrillo, J. M., Wang, Y., Sumpter, B. G., Huang, G.-R., Do, C. & Chen, W.-R. (2022). *Commun. Phys.* **5**, 46.
- Chen, S.-H. (1986). *Annu. Rev. Phys. Chem.* **37**, 351–399.
- Chu, B. & Hsiao, B. S. (2001). *Chem. Rev.* **101**, 1727–1762.
- Debye, P. (1947). *J. Phys. Chem.* **51**, 18–32.
- De Gennes, P. (1979). *Scaling concepts in polymer physics*. Cornell University Press.
- De Gennes, P.-G. (1990). *Introduction to polymer dynamics*. Cambridge University Press Archive.
- Ding, L., Tung, C.-H., Cao, Z., Ye, Z., Gu, X., Xia, Y., Chen, W.-R. & Do, C. (2025a). *Digital Discov.* **4**, 1570–1577.
- Ding, L., Tung, C. H., Sumpter, B. G., Chen, W. R. & Do, C. (2024). *J. Chem. Theory Comput.* **20**, 10697–10702.
- Ding, L., Tung, C. H., Sumpter, B. G., Chen, W. R. & Do, C. (2025b). *J. Chem. Theory Comput.* **21**, 4176–4182.
- Feng, C. & Huang, X. (2018). *Acc. Chem. Res.* **51**, 2314–2323.
- Huang, G.-R., Porcar, L., Murphy, R. P., Shinohara, Y., Wang, Y., Carrillo, J.-M., Sumpter, B. G., Tung, C.-H., Ding, L., Do, C. & Chen, W.-R. (2025). *J. Appl. Cryst.* **58**, 637–658.
- Huang, G.-R., Tung, C.-H., Porcar, L., Wang, Y., Shinohara, Y., Do, C. & Chen, W.-R. (2023). *Macromolecules* **56**, 6436–6443.
- Kholodenko, A. (1993). *Macromolecules* **26**, 4179–4183.
- Krauth, W. (2006). *Statistical mechanics: algorithms and computations*, Oxford master series in statistical, computational, and theoretical physics, Vol. 13. Oxford University Press.
- Kremer, K. & Grest, G. S. (1990). *J. Chem. Phys.* **92**, 5057–5086.
- Li, R. & Wang, J. (2016). *Polymers* **8**, 328.
- Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y. & Tegmark, M. (2024). *arXiv*, 2404.19756.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT Press.
- Murphy, R. P., Riedel, Z. W., Nakatani, M. A., Salipante, P. F., Weston, J. S., Hudson, S. D. & Weigandt, K. M. (2020). *Soft Matter* **16**, 6285–6293.
- Pedersen, J. S. & Schurtenberger, P. (1996). *Macromolecules* **29**, 7602–7612.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011). *J. Mach. Learn. Res.* **12**, 2825–2830.
- Ren, J. M., McKenzie, T. G., Fu, Q., Wong, E. H., Xu, J., An, Z., Shanmugam, S., Davis, T. P., Boyer, C. & Qiao, G. G. (2016). *Chem. Rev.* **116**, 6743–6836.
- Robbins, M. O., Kremer, K. & Grest, G. S. (1988). *J. Chem. Phys.* **88**, 3286–3312.
- Schroeder, C. M., Teixeira, R. E., Shaqfeh, E. S. & Chu, S. (2005). *Macromolecules* **38**, 1967–1978.
- Shibayama, M. (2011). *Polym. J.* **43**, 18–34.
- Smith, D. E., Babcock, H. P. & Chu, S. (1999). *Science* **283**, 1724–1727.
- Sperling, L. H. (2015). *Introduction to physical polymer science*. John Wiley & Sons.
- Strang, G. (2022). *Introduction to linear algebra*. SIAM.
- Tung, C.-H., Chang, S.-Y., Chang, M.-C., Carrillo, J.-M., Sumpter, B. G., Do, C. & Chen, W.-R. (2023). *Carbon Trends* **10**, 100252.
- Tung, C. H., Chang, S. Y., Chen, H. L., Wang, Y., Hong, K., Carrillo, J. M., Sumpter, B. G., Shinohara, Y., Do, C. & Chen, W. R. (2022). *J. Chem. Phys.* **156**, 131101.
- Tung, C.-H., Chen, M.-Z., Chen, H.-L., Huang, G.-R., Porcar, L., Chang, M.-C., Carrillo, J.-M., Wang, Y., Sumpter, B. G., Shinohara, Y., Do, C. & Chen, W.-R. (2024a). *J. Appl. Cryst.* **57**, 1047–1058.
- Tung, C. H., Ding, L., Chang, M. C., Huang, G. R., Porcar, L., Wang, Y., Carrillo, J. Y., Sumpter, B. G., Shinohara, Y., Do, C. & Chen, W. R. (2025). *J. Chem. Phys.* **162**, 074106.
- Tung, C. H., Hsiao, Y. J., Chen, H. L., Huang, G. R., Porcar, L., Chang, M. C., Carrillo, J. M., Wang, Y., Sumpter, B. G., Shinohara, Y., Taylor, J., Do, C. & Chen, W. R. (2024b). *J. Colloid Interface Sci.* **659**, 739–750.
- Wang, Y. & Wang, S.-Q. (2011). *Macromolecules* **44**, 5427–5435.
- Williams, C. K. & Rasmussen, C. E. (2006). *Gaussian processes for machine learning*, Adaptive computation and machine learning series, Vol. 2, No. 3. MIT Press.
- Zhu, M. & Ghodsi, A. (2006). *Comput. Stat. Data Anal.* **51**, 918–930.