Reassessing the transport properties of fluids: A symbolic regression approach

Dimitrios Angelis,* Filippos Sofos,† and Theodoros E. Karakasidis‡
Condensed Matter Physics Laboratory, Department of Physics, University of Thessaly, Lamia 35100, Greece

(Received 14 September 2023; accepted 4 January 2024; published 25 January 2024)

The viscosity and thermal conductivity coefficients of the Lennard-Jones fluid are extracted through symbolic regression (SR) techniques from data derived from simulations at the atomic scale. This data-oriented approach provides closed form relations that achieve fine accuracy when compared to well-established theoretical, empirical, or approximate equations, fully transparent, with small complexity and high interpretability. The novelty is further outlined by suggesting analytical expressions for estimating fluid transport properties across the whole phase space, from a dilute gas to a dense liquid, by considering only two macroscopic properties (density and temperature). In such expressions, the underlying physical mechanisms are reflected, while, at the same time, it can be a computationally efficient alternative to costly in time and size first principle and/or molecular dynamics simulations.

DOI: 10.1103/PhysRevE.109.015105

I. INTRODUCTION

As thermophysical property data is constantly becoming available, mainly through atomic-scale simulations, and some well-posed experiments, reevaluation methods, and techniques are proposed to produce better correlations between properties of interest and provide insight to the physics of the underlying physical system. Empirical or approximate relations have been the basic theoretical tool to lead the establishment of related theories during the past decades while in the last few years, machine learning (ML) based methods have emerged. In the era of big data, it would be highly profitable to investigate the possibility of unraveling complex physical processes through statistical data analysis, as long as the proposed methods are set on concrete ground and focus on enhancing and binding current knowledge into new directions.

Atomic scale simulations, such as molecular dynamics (MD), have proven to be an invaluable choice to calculate fundamental fluid properties (i.e., density, specific heat, compressibility, and more) through interatomic potentials, such as the Lennard-Jones potential. On the other hand, fluid transport properties such as shear viscosity, and thermal conductivity, require computationally intensive and timely MD simulations, and suggested models were focused on suggesting an accurate framework able to reproduce the available experimental results [1–10]. Before MD simulations take the leading role in property extraction, the Enskog transport theory, which was based on the hard sphere (HS) fluid model, had been widely incorporated, especially for the investigation of low-density fluids [11]. For denser fluids, more complex approaches have been introduced, such as the excess-entropy method [12]. The excess-entropy has been found to control liquid equilibrium transport properties for simple liquids which closely resemble the behavior of HS models, while other recent efforts have investigated this connection [13], improving the accuracy and scaling of the initial theory [14].

Other approaches include kinetic theory principles, where collision integrals have to be calculated [15–17], and the inversion method [18–20]. The field of application in real fluids includes ionic liquids [21,22], gases [23], thermoelectric analysis [24,25], polar and nonpolar fluids [26,27], and more [28,29]. One can find more information in several well-written reviews [30,31].

However, the kinetic solution for dilute monatomic gases cannot be applied for polyatomic molecules [18]. In such cases, molecules undergo several inelastic collisions and therefore kinetic energy is not conserved (on the opposite, mass and momentum are). As a result, that theory is only applicable to low pressure, high temperature, nonpolar gases [23]. To overcome this obstacle, several empirically correlated density-dependent functions are being introduced into the existing kinetic gas theory, in order to extend calculations of transport properties at higher densities [26]. Similarly, methods (e.g., [32]) are bound to liquid phases [27], or nanofluid procedures are hindered by the type of nanofluid, the size of nanoparticle, and volume fractions [30].

It is well known that the accuracy of MD simulations rely on the quality of the interatomic potential employed. On the one hand, first principle MD simulation attempts have been reported, but on the other hand, the huge computational demands hamper their application to common materials [33]. In [34], MD simulations based on the density functional theory (DFT) framework has given accurate transport property values for Ar and water, that are close to literature results. Another methodology within the framework of Maxwell viscoelastic theory has been presented for local shear viscosity calculations in nanopores [35]. Recently, an integrated computational framework, FeOs [36], has been proposed, which can be used to compute various properties of

*dimangelis@uth.gr
†Corresponding author: fsofos@uth.gr
‡thkarak@uth.gr

materials (pure or mixtures), with a basis on equations of state and DFT.

A conclusive remark on all these methods is that, notwithstanding their wide applicability and proven accuracy with experimental results, their restriction to a certain extent of experimental conditions, their computational burden, and the fact that the proposed relations fail to generalize over a fluid phase transition, are still open issues. Next generation approaches employed in physical sciences are data driven and focus on data-derived knowledge through ML [37]. The introduction of ML for material properties calculations has provided increased accuracy, followed by reduced computational cost [38–40], by exploiting the wealth of available data in the literature. In more advanced schemes, an ML framework running in parallel to the simulation can be exploited in order to train a macroscopic model with microscopic data [41] in different time and length scales [42]. In physical sciences, it would be highly desirable to transform the available data into explainable and interpretable mathematical relations [43].

Genetic programming-based (GP) approaches, which code natural species evolution mechanisms into computational tools, can have a primitive role toward this direction. More specifically, a subfield of GP, symbolic regression (SR), is capable of deriving analytical expressions that fit in a particular dataset, only by focusing on data behavior [44] and without prior knowledge of the investigated system [45]. Overpassing the traditional computational and statistical approaches [46–53], SR provides a different outlook to reveal the physical data behavior and many recent works have shown that it can be successfully applied to diffusion coefficient prediction [54–56], ionic conductivity [57], lattice thermal conductivity [58], and viscosity [59,60], to mention a few.

In the upcoming sections, a ML procedure is implemented for viscosity and thermal conductivity analysis. The aim is to obtain symbolic expressions (SE) for shear viscosity and thermal conductivity of the Lennard-Jones (LJ) fluid as a function of two macroscopic properties, fluid density and temperature. The investigation is spread across all fluid states available in the MD datasets, from dilute gases to dense liquids. At first, several conventional approaches that are widely used for shear viscosity and thermal conductivity analysis and prediction are briefly introduced. Second, we focus on the adopted ML methodology, where a preprocessing data analysis is performed to reveal statistical correlations between input features and the properties of interest, followed by the SR algorithm description. Resulting ML-derived equations are presented and discussed on their physical correspondence. It is found that the proposed equations for shear viscosity and thermal conductivity achieve sufficient accuracy and it is concluded that SR could be embedded in transport properties prediction, either in interpolation or extrapolation, bypassing simulation techniques wherever possible.

II. MATHEMATICAL MODELS AND METHODS

A. The transport properties of fluids

In general, transport properties such as viscosity and thermal conductivity can be described at a molecular point of view by statistical mechanics equations that account for intermolecular forces [61]. These quantities have gained attention due to their ability to control the transfer of momentum (viscosity) and thermal energy (thermal conductivity). Moreover, as fluid transforms from dilute gas into dense liquid, an increase at the values of transport properties for constant temperature has been reported, which could be a hint that different mechanisms might govern these phases [62]. Hence, decoding this behavior is of particular interest.

It is a fact that several mathematical (empirical and/or approximate) relations have emerged, but the inherent limitation that bounds them into a certain density interval or the fact that they contain adjustable parameters prone to experimental measurement uncertainty, makes them of limited practical applicability. For instance, viscosity and thermal conductivity can be described by the Chapman-Enskog solution of the Boltzmann kinetic theory, which accounts for the binary interactions of the molecules and correlates the former properties into various collision integrals ($\Omega^{(l,s)}$). These collision integrals are defined by [18]

$$\Omega^{(l,s)} = [(s+1)!(k_B T)^{s+2}]^{-1} \int_0^\infty Q^{(l)}(E) e^{-E/k_B T} E^{s+1}\, dE, \tag{1}$$

$$Q^{(l)}(E) = 2\pi \left[1 - \left(1 + \frac{-1^l}{2(l+1)}\right)\right]^{-1} \int_0^\infty (1 - \cos^l\theta) b\, db, \tag{2}$$

$$\theta(b,E) = \pi - 2b \int_0^\infty \left[\frac{1-b^2}{r^2} - \frac{V(r)}{E}\right]^{-1/2} \frac{1}{r^2}\, dr, \tag{3}$$

with k_B being the Boltzmann constant, T is temperature, l and s are weighting factors linked to the transport mechanism due to collisions, E being the relative energy of colliding molecule, $Q^{(l)}(E)$ is a transport collision integral, θ is the scattering angle between two colliding molecules with energy E and b impact parameter at the gas temperature, r is the distance of two molecules, and $V(r)$ is the intermolecular pair potential function. However, this approach exhibits deviations when implemented at higher densities.

The transport properties of fluids are usually extracted via MD simulations. By screening the particle trajectories, velocities, and calculating interatomic forces during simulations, it is possible to derive time-dependent properties at equilibrium or nonequilibrium conditions. These approaches, highly rely on the accuracy of the chosen interatomic potential. Some decades ago, simple numerical approaches for dilute gases have depended on the HS potential [63]

$$\Phi_{\mathrm{HS}} = \begin{cases} \infty, & r < \sigma \\ 0, & r > \sigma \end{cases}, \tag{4}$$

where Φ_{HS} is the HS potential of two spheres with σ diameter at r distance. For this potential, it is hypothesized that all particles are finite sized and the collisions are strictly elastic. Transport coefficient relations produced from this potential are accurate at gas phases but weaken while approaching dense liquid states. This is mainly because the above interatomic potential is perhaps too simple and cannot account for more complex interactions. On the contrary, the widely used LJ potential has been found to perform better on the

calculation of the transport coefficients of real fluids, providing results with increased accuracy (e.g., several noble gases [64]). The 12–6 LJ potential Φ_{LJ} is given by [65]

$$\Phi_{LJ} = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{5}$$

where r is the distance between particles, σ is the interatomic separation where the potential energy is zero, and ϵ is the potential well depth. LJ potential encompasses variables such as density (ρ), temperature (T), distance (r), viscosity (η), thermal conductivity (λ), and time (t) with a special trait that they appear dimensionless (oftentimes symbolized by an asterisk in superscript). These dimensionless variables are merely a product of simulation features, scaled by ϵ and σ and defined by the following equations [66,67]:

$$\rho^* = \rho\sigma^3, \quad T^* = \frac{k_B T}{\epsilon}, \quad r^* = \frac{r}{\sigma},$$
$$\eta^* = \eta\frac{\sigma^2}{\sqrt{m\epsilon}}, \quad \lambda^* = \lambda\frac{\sigma^2}{k_B}\left(\frac{m}{\epsilon}\right)^{1/2}, \quad t^* = \frac{t\sqrt{\epsilon/m}}{\sigma}, \tag{6}$$

with k_B being the Boltzmann constant, σ and ϵ are the LJ parameters, N is the number of particles, m is the mass of the atom in pure system, and V is the volume of the simulation box.

1. Shear viscosity

During gas phase, transport of momentum takes place by molecular collision, whereas at liquid phase, molecules interact within a more dense environment, and the momentum transfer is achieved by advancing forces between them [68]. The difference between these procedures leads to the assumption that different approaches might be more precise in order to estimate the viscosity coefficient. In fact, it is stated that gas theories are more advanced when compared to liquid approaches, where the complexity makes it difficult to produce a solid basis [68].

Starting from low densities, viscosity of a dilute monatomic gas can be expressed in the first Enskog approximation as [1]

$$\eta = \frac{5}{16}\left(\frac{mk_B T}{\pi}\right)^{1/2}\frac{1}{\sigma^2\Omega^{(2,2)*}}, \tag{7}$$

with η being the viscosity, m is the molecular mass, T is the temperature, k_B is the Boltzmann constant, and σ is the interatomic separation. $\Omega^{(2,2)*}$ is the collision integral of Eq. (1) where $l, s = 2$ and scaled by π and σ, via [18]

$$\Omega^{(l,s)*} = \frac{\Omega^{(l,s)}}{\pi\sigma^2}. \tag{8}$$

Moreover, the temperature dependence is different for each interatomic potential chosen, while what should also be noted is that a correction multiplicative factor (f_η) is occasionally included in Eq. (7); however, most of the times, it is considered as $f_\eta = 1$ [18].

Alternatively, MD simulations have a prominent advantage over other theoretical approaches. For a system at equilibrium, shear viscosity (and thermal conductivity, as will be shown next) can be calculated via the Green-Kubo (GK) formulation [69]. The shear viscosity for a pure fluid is obtained by

$$\eta = \frac{1}{Vk_B T}\int_0^\infty dt\langle J_p^{xy}(t)\cdot J_p^{xy}(0)\rangle, \tag{9}$$

with J_p^{xy} being the off-diagonal elements of the microscopic stress tensor, described by

$$J_p^{xy} = \sum_{i=1}^{N} m_i v_i^x v_i^y - \sum_{i=1}^{N}\sum_{j>1}^{N} r_{ij}^x \frac{\partial u(r_{ij})}{\partial r_{ij}^y}, \tag{10}$$

where $u(r_{ij})$ is the LJ potential of the interacting atoms i and j, $r_{i,j}$ is the distance between these atoms, and v_i^j is the j component (whereas $j = x, y, z$) of the velocity that corresponds to the i atom.

At nonequilibrium conditions, GK can also be incorporated, however, care has to be taken in order to stay in the linear regime, not far from equilibrium [70]. In the presence of strain rates, where fluid layers (in the y direction) tend to move toward each other (in the x direction), nonequilibrium MD (NEMD) methods are exploited [71], and the shear viscosity is obtained as a profile in y axis from

$$\eta(y) = -\frac{P_{xy}}{\gamma(y)}, \tag{11}$$

where γ is the strain rate and P_{xy} is the time-averaged, off-diagonal component of the microscopic stress tensor. For more details on the equation, the reader refers to [72]. In this paper we focus only on bulk, equilibrium systems.

It has been observed that the viscosity increases in denser fluids at constant temperature [62]. Moreover, at gas phase, an increase of viscosity values is expected when temperature increases. On the other hand, at liquid phase, viscosity should decrease at higher temperature. This contradicting behavior can be anticipated with the introduction of density fluctuation theory (DFT). According to DFT, shear viscosity is in fact a product of two different parts, the kinetic and the potential [73],

$$\eta = \eta_k + \eta_v, \tag{12}$$

where the kinetic part is given by Eq. (7), i.e., the Enskog approximation for viscosity of dilute gases and the potential part is equal to

$$\eta_v = \frac{\rho^2\omega(\rho, T)}{6D}, \tag{13}$$

with $\omega(\rho, T)$ given by

$$\omega(\rho, T) = \frac{2\pi}{15}\int_0^\infty dr\, r^5\, V'(r)g(r, \rho, T)\theta(\xi - r), \tag{14}$$

where $V(r)$ is the LJ interaction potential energy, $g(r, \rho, T)$ is the dynamic radial distribution function, and θ characterizes the nature of density fluctuations over the ξ range [73,74]. Conversely, by following the general behavior of of Eq. (12), it is stated that the constitutive parts present a diametrically opposite way of acting at gases and liquids regarding temperature and density, which basically describes the viscosity transition of gases into liquids. In addition, as phase shifts from dilute gas into dense liquid and vice versa, each equation term may surpass the values of the other at the corresponding regimes [73].

TABLE I. Dataset description.

Dataset (assembled from Ref.)	No. of observations	Parameter value range		
		Density (ρ^*)	Temperature (T^*)	Transport property (η^*, λ^*)
Shear viscosity (η^*) [66]	343	0.005–1.275	0.698757–6.004	0.04947–11.06
Thermal conductivity (λ^*) [67]	134	0.05–0.95	0.6–4.0	0.6336–11.48

2. *Thermal conductivity*

Following a similar approach to viscosity investigation, an expression to calculate the thermal conductivity of a dilute monatomic gas can be obtained by the first Enskog approximation [1]:

$$\lambda = \frac{75}{64}\left(\frac{k_B^3 T}{\pi m}\right)^{1/2}\frac{1}{\sigma^2\Omega^{(2,2)*}}, \tag{15}$$

with λ being the thermal conductivity, m is the molecular mass, T is the temperature, k_B is Boltzmann's constant, and $\Omega^{(2,2)*}$ is the reduced collision integral of Eq. (8).

Thermal conductivity can also be divided in two distinct terms, the kinetic and potential part [75], as

$$\lambda = \lambda_k + \lambda_v. \tag{16}$$

The kinetic contribution, λ_k, is given by the Enskog approximation of thermal conductivity for dilute gases [Eq. (15)], while the potential contribution, λ_v, is given by

$$\lambda_v = \frac{\rho^2\chi(\rho,T)}{6D}, \tag{17}$$

where $\chi(\rho,T)$ is defined by

$$\chi(\rho,T) = \frac{4\pi k_B\kappa^*}{m}\int_0^\infty dr_{12}r_{12}^4\left[\frac{r_{12}V'(r_{12})}{5} - V(r_{12})\right] \times g(r_{12};\rho)\theta(r_{12}-\xi). \tag{18}$$

More details can be found in [75].

The statistical mechanics GK formula for thermal conductivity is given by

$$\lambda = \frac{1}{Vk_BT^2}\int_0^\infty dt\left\langle J_q^x(t)\cdot J_q^x(0)\right\rangle, \tag{19}$$

with $\mathbf{J}_q$ being the microscopic heat flow which is given by

$$\mathbf{J}_q = \frac{1}{2}\sum_{i=1}^{N} m_i(v_i)^2\mathbf{v}_i - \sum_{i=1}^{N}\sum_{j>1}^{N}\left[\mathbf{r}_{ij}:\frac{\partial u(r_{ij})}{\partial \mathbf{r}_{ij}} - \mathbf{I}\cdot u(r_{ij})\right]\cdot\mathbf{v}_i, \tag{20}$$

where $\mathbf{v}_i$ is the velocity vector of i atom and $\mathbf{I}$ is the unitary matrix.

In general, thermal conductivity for simple organic liquids increases as the fluid shifts from dilute gas into dense liquid, where typical values are between two orders of magnitude larger. At gas phase with low pressure and high temperature, an increase of thermal conductivity is expected, while at liquid phase, as temperature increases, the value of thermal conductivity decreases [62].

B. Data analysis

Viscosity and thermal conductivity simulation data are taken from literature sources, originated from MD simulations of the Lennard-Jones fluid. Table I presents some basic data features. To feed the SR algorithm, the datasets are randomly split into a training and a test (validation) dataset via a partitioning factor of 80% and 20%, respectively.

As shown in Fig. 1(a), data for shear viscosity expand to a wide region of the (ρ^*-T^*) phase space, incorporating gas, liquid, dense-liquid, and supercritical states, while the available data points for thermal conductivity are mainly gathered at the liquid/dense liquid region. The effect of fluid density for T^* = constant is depicted in Fig. 1(d). Here we notice an increase of viscosity in denser fluids, and this behavior is likely to be approached by some kind of exponential or power function. On the other hand, density and thermal conductivity diagrams exhibit a more direct increase of thermal conductivity; however, this is highly more likely to be formed of a parabolic or maybe an exponential function, rather than a linear. Conversely, temperature and transport property plots [Fig. 1(c)] indicate a connection that resembles in a more or less linear proportional function. Finally, from the correlation tables in Fig. 1(b), viscosity appears highly positive correlated to density values and low correlated to temperature. In case of thermal conductivity, the correlation plot reveals high positive correlation to temperature, while density and thermal conductivity are not correlated at all.

C. Symbolic regression

Symbolic regression derives a closed form SE which describes, to a certain extent, the underlying mechanisms inside a collection of data [45]. Other ML-based methods such as neural networks (NNs), are generally employed to derive numerical predictions from data. However, these are based on "black-box" models, that do not provide a physical interpretation of the investigated system and this might lead to an abstract mapping of the system's behavior [78]. Alternatively, SR is able to model the system's behavior through an analytical equation. Although not producing new theory, this data-driven approach can provide an equation to be used at hand, making it more suitable for physical-related tasks: explainable, interpretable, and generelizable.

In general, the supersets of SR, evolutionary algorithms (EA) [79] and genetic programming (GP) [80], create a large search space (population) filled by random solutions (individuals) and locate those that achieve higher performance (in terms of a loss function). The final step includes a random recombination process into a new set of equations via crossover. This is an iterative process that creates as many populations as selected by each user or when the loss functions becomes

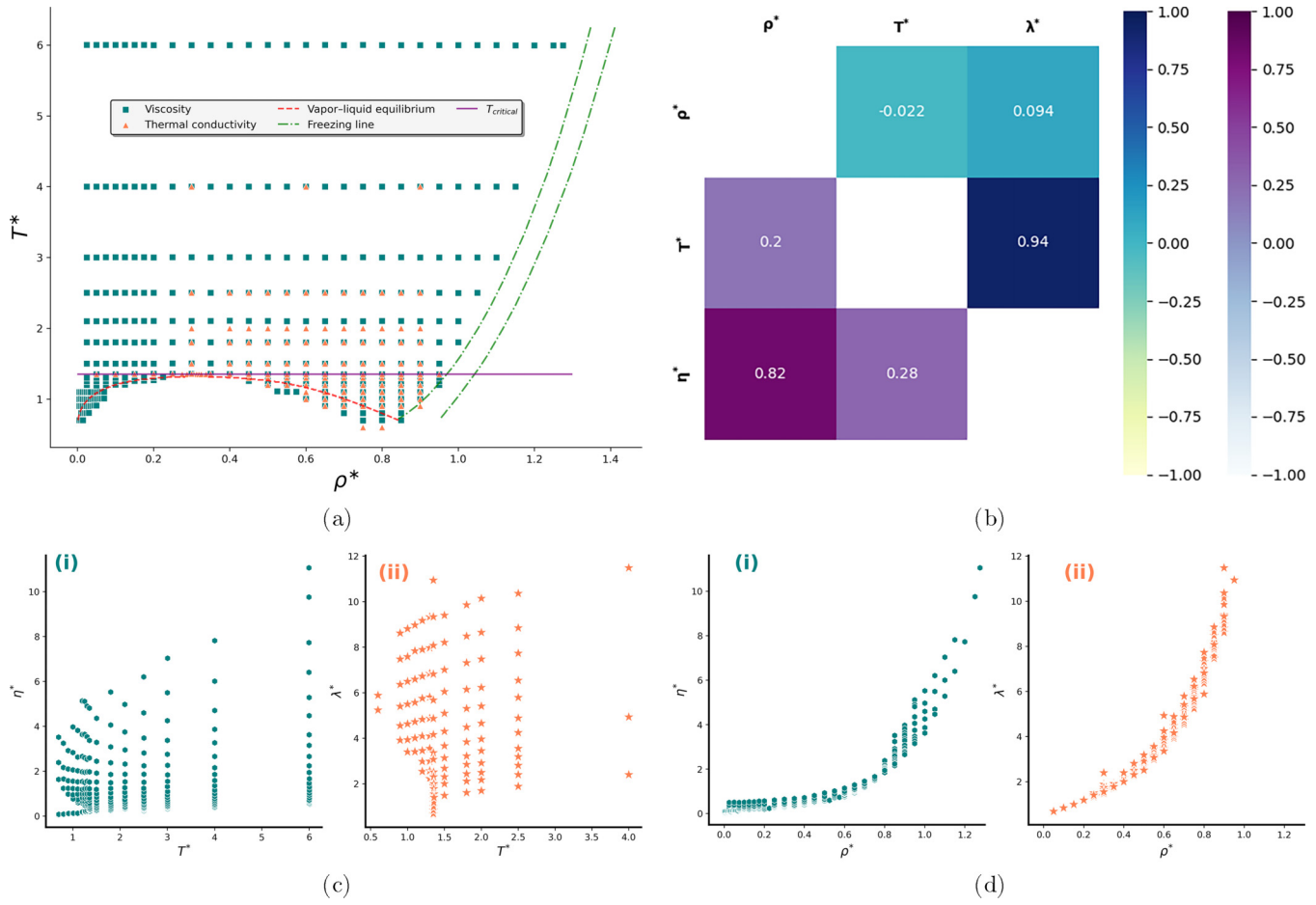


FIG. 1. Statistical analysis for shear viscosity and thermal conductivity. (a) Phase diagram. Dashed lines denote the critical temperature, vapor-liquid equilibrium (VLE) [76], and freezing line [77]. (b) Mixed correlation matrix, where the upper triangle is for λ^* and the lower for η^*. (c) Transport properties vs temperature. (d) Transport properties vs density, where (i) corresponds to viscosity and (ii) to thermal conductivity.

smaller than a given value. Random mutations of these solutions are also being introduced to minimize the possibility of overfitting [81]. A schematic diagram of the above procedure is illustrated in Fig. 2. More details on the method can be found in the recent review of Ref. [43].

For the algorithmic implementation, the SR-proposed framework here is based on the SR library package included in the open python-julia software PySR [82]. The conquering equations are selected based on a loss function metric, along with an estimation of the equation complexity metric. The loss function applies to the mean square error [MSE, Eq. (21)],

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2, \tag{21}$$

while the number of nodes that correspond to a treelike equation accounts for the level of complexity. Therefore, the total score for each proposed equation lies in the pareto front [83].

The generated expressions appear in a tree structure form, whilst the selection of the best performing equations is twofold. First, a loss function determines the degree of fitness these individuals achieve over the dataset. In our case, we have MSE as the loss function. Second, complexity values are assigned at each individual, which corresponds to the number of their tree nodes, and the total score is a product of two. Moreover, individuals are being reproduced by the typical evolutionary mutation operators (e.g., crossover, mutation) and after a number of iterations, each population achieves a high score that lies at the pareto front [83]. Each population develops autonomous knowledge, however, equations that fit to a specific problem are constantly appearing in the results pool. In the end, physical knowledge and intuition,

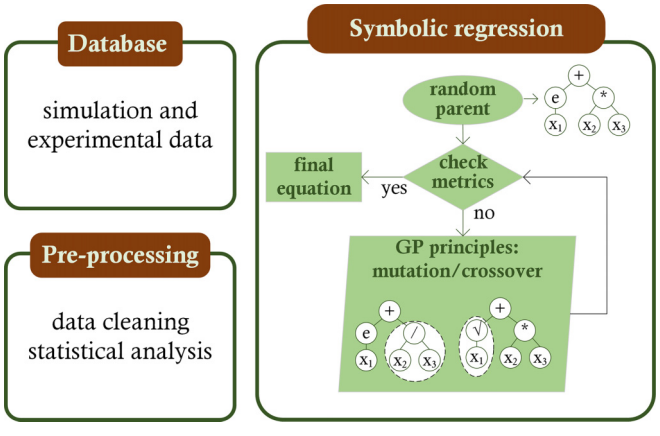


FIG. 2. Symbolic regression approach.

TABLE II. Symbolic expressions measures of accuracy.

		Training dataset (80%)				Test dataset (20%)			
Equation	*Compl.*	R^2	*MSE*	*RMSE*	*MAE*	R^2	*MSE*	*RMSE*	*MAE*
$\eta_1^{*\text{symb}}$ (22)	15	0.99	0.02	0.14	0.11	0.99	0.02	0.14	0.10
$\eta_2^{*\text{symb}}$ (23)	15	0.99	0.03	0.17	0.11	0.99	0.03	0.17	0.12
$\lambda_1^{*\text{symb}}$ (24)	17	1.0	0.02	0.14	0.1	1.0	0.02	0.14	0.1
$\lambda_2^{*\text{symb}}$ (25)	9	0.99	0.04	0.20	0.13	0.99	0.04	0.20	0.14

in a manual manner, are exploited to select these expressions that achieved low error metrics, low complexity, and, most importantly, can be explained in physical terms for the problem investigated.

III. RESULTS

A. Symbolic regression equations

The SR framework, as presented in Fig. 2, is implemented in order to acquire viscosity and thermal conductivity coefficients as functions of density and temperature. After a great number of parallel runs, a pool of candidate equations have been extracted. Attention has been drawn on those equations that keep appearing in the output, even though the algorithm starts by considering random initial parents every time. This case of equation reappearances in such a random process could be an indication of capturing some kind of physical behavior inside data. Finally, we ended up in two equations for viscosity, Eqs. (22) and (23), and two for thermal conductivity, Eqs. (24) and (25), as

$$\eta_1^{*\text{symb}} = a_1\sqrt{T^*} + a_2\rho^{*4}\left(a_3 + \frac{a_4\rho^*}{\sqrt{T^*}}\right)^2, \quad T^* \neq 0, \tag{22}$$

$$\eta_2^{*\text{symb}} = \sqrt{\rho^*} + \rho^*(b_1 - \log(T^*))(\rho^{*4} - b_2), \quad T^* \neq 0, \tag{23}$$

$$\lambda_1^{*\text{symb}} = c_1\sqrt{T^*} + \rho^{*2}(T^* + \rho^* + e^{2\rho^*} + c_2), \tag{24}$$

$$\lambda_2^{*\text{symb}} = e^{(\rho^*+d_1)^2} - \frac{d_2}{T^*}, \quad T^* \neq 0, \tag{25}$$

where a_i, b_j, c_k, d_l $(i = 1-4,\ j = 1,2,\ k = 1,2,\ l = 1,2)$ coefficients have been approximately calculated as $a_1 = 0.21, a_2 = 2.06, a_3 = 1, a_4 = 0.7, b_1 = 4.84, b_2 = 0.09, c_1 = 0.9, c_2 = 1.9, d_1 = 0.635, and d_2 = 1.28$.

It should be noted that $\eta_1^{*\text{symb}}$ [Eq. (22)] and $\eta_2^{*\text{symb}}$ [Eq. (23)] present a high order polynomial dependence on density. This is somewhat expected, if we take into account the behavior presented in Fig. 1(d), where viscosity presents a polynomial increase over density increase. Temperature dependence appears in $\eta_1^{*\text{symb}}$ as square root, while it is more complicated in $\eta_2^{*\text{symb}}$, where a temperature logarithmic term is drawn.

In case of thermal conductivity, $\lambda_1^{*\text{symb}}$ [Eq. (24)] denotes a square root correlation on temperature and, in parallel, a more complex interconnection with density, with square, cubic, and exponential dependencies identified. In contrast, $\lambda_2^{*\text{symb}}$ [Eq. (25)] is simpler, inversely proportional, and negatively affected by temperature, while density has some kind of exponential relation.

The MD simulations dataset exploited for equation extraction contains measurements across every fluid phase (gas, liquid, dense liquid, and supercritical). Therefore, we expect that the proposed equations account for fluid behavior across all states, along with critical regions and phase transitions from gas to liquid. However, we have to bear in mind that SR is not a process of suggesting new physics; we consider it as a computational tool that may be able to reveal physical trends only from data, where possible, without former assumption on its behavior. Next, the suggested equation accuracy is analyzed.

B. Statistical post processing

To argue on the proposed equations accuracy, several statistical measures have been calculated, such as the coefficient of determination R^2, *MSE*, the root MSE (*RMSE*), and the mean absolute error (*MAE*). To account for overfitting, each equation is rerun at a range of 0–100 different random initial states. The average value of the statistical measures, along with equation complexity (Compl.) (analog to the number of tree nodes in each equation) are presented in Table II.

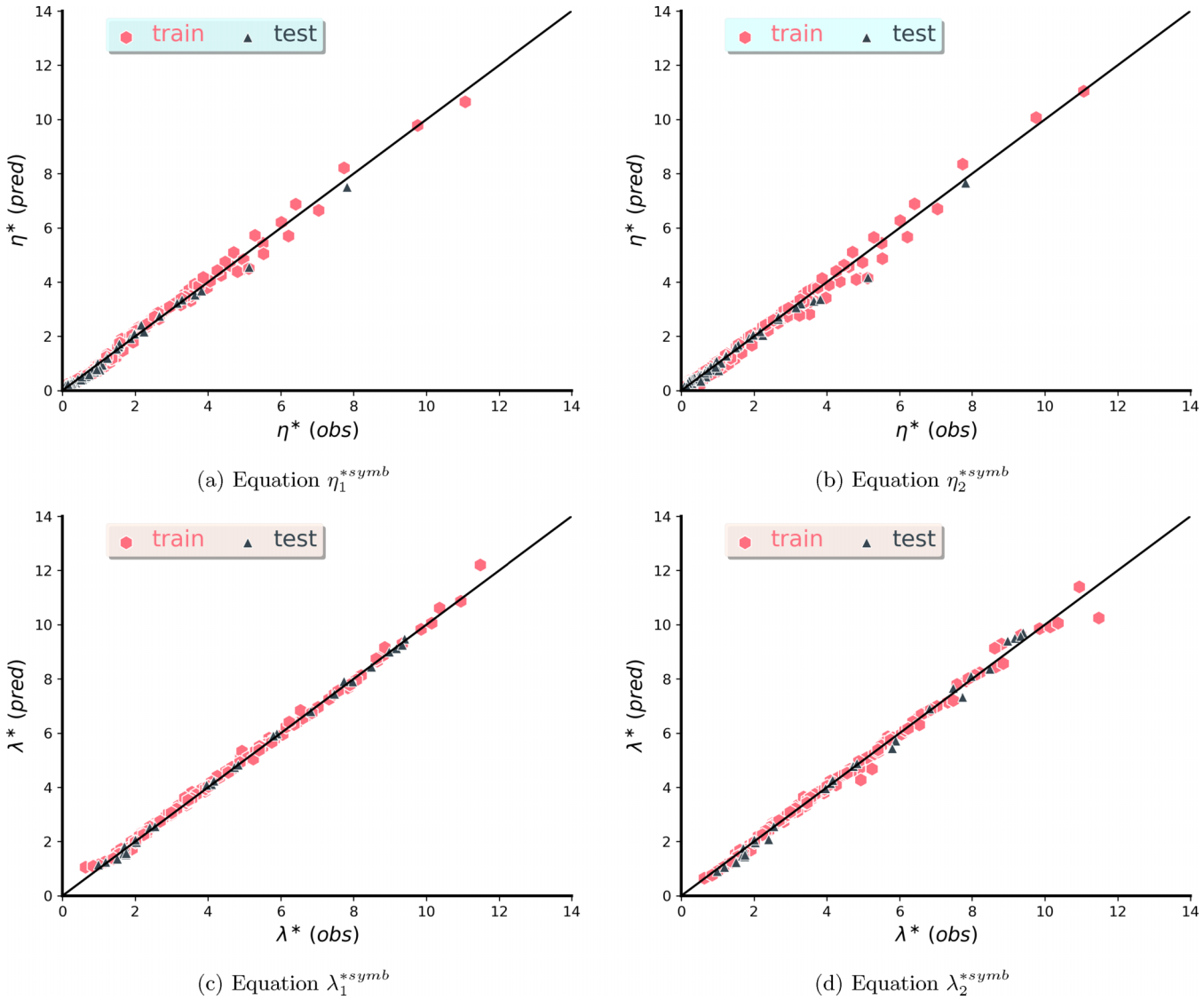
The identity plots for viscosity equations, $\eta_1^{*\text{symb}}$ and $\eta_2^{*\text{symb}}$, in Fig. 3 present an excellent fit between simulation and predicted values. This is also verified by low error metrics in Table II. Another point worth mentioning is the measure similarity between the training and the test dataset, which is an indication of no overfitting. Low complexity of both equations is, further, an asset. Nevertheless, some deviations appear at intermediate viscosity values (mostly for $3.0 \leqslant \eta^* \leqslant 8.0$).

Equally noteworthy is the fact that thermal conductivity identity plots suggest that both equations fit well on simulation data. Moreover, despite the significant difference on complexity values, accuracy measures remain comparable (see Table II). No overfitting occurs here as well.

IV. DISCUSSION

Apart from fine fitting of the proposed equations on MD simulation data, a physical connection to the hidden nature of the transport properties of fluids is further needed. We start the investigation by considering $\rho^* =$ constant and plot viscosity and thermal conductivity values vs temperature at different fluid regions (gas, liquid, dense liquid) in Fig. 4.

The y-axis scale is different in each fluid state, since viscosity is higher at high-dense liquids compared to liquids



(a) Equation η_1^{*symb}

(b) Equation η_2^{*symb}

(c) Equation λ_1^{*symb}

(d) Equation λ_2^{*symb}

FIG. 3. Identity plots of symbolic expressions. The 45° line denotes perfect fit between equation and available data.

and, even more, to gases. An increase of viscosity values at gas phase for higher temperature is expected, and this is observed for both equations (at least for values below the critical point), in Figs. 4(a) and 4(b). When at liquid state, a fluid confronts decrease of viscosity as temperature increases. This premise is also fulfilled for both proposed equations. Last but not least, it seems that both equations perform relatively well at liquid and dense liquid regimes, since deviations are minimum. On the other hand, different results are obtained for the gas state, where both $\eta_2^{*\text{symb}}$ and $\eta_1^{*\text{symb}}$ fail to capture simulation data behavior.

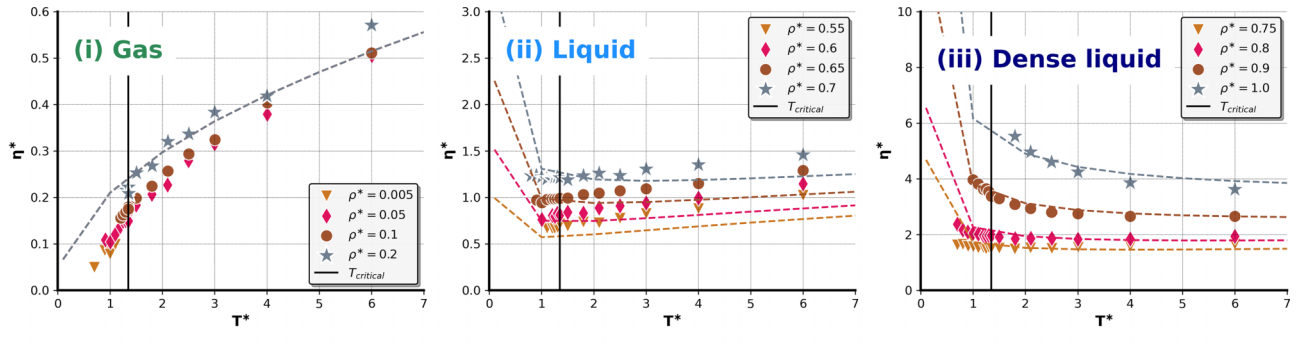
Conversely, it is not straightforward to draw physical conclusions from the suggested thermal conductivity equations at gas states [Figs. 4(c) and 4(d)], where the implied dataset does not contain ample measurements. The physical trend of thermal conductivities increase as the fluid transforms from gas to liquid, and dense liquid is followed by the proposed equations and the available data. What is also evident is a continuous increase of λ at constant density values when temperature increases, as is also shown in the density-thermal conductivity diagram in Fig. 1(d). On the opposite, $\lambda_2^{*\text{symb}}$ displays a peculiar behavior for temperature values below the critical value, as a dramatic decrease is noticed for $T^* < 1.0$. In addition, the fact that the proposed interdependence between thermal conductivity and temperature does not adhere to the statistical preprocessing analysis shown, we believe that $\lambda_2^{*\text{symb}}$ might not be a potential thermal conductivity relation.

The SR-extracted equations are further investigated on the limits of density and temperature values. At dilute gas conditions, we practically reach the zero density limit. An inherent advantage is that every equation is defined at zero density limit. Therefore, as density approaches zero, each remaining expression $\eta_1^{*\text{symb}}$, $\eta_2^{*\text{symb}}$, $\lambda_1^{*\text{symb}}$ becomes $\eta_{1,\text{dilute}}^{*\text{symb}}$, $\eta_{2,dilute}^{*\text{symb}}$, $\lambda_{1,\text{dilute}}^{*\text{symb}}$, accordingly, and equations transform to

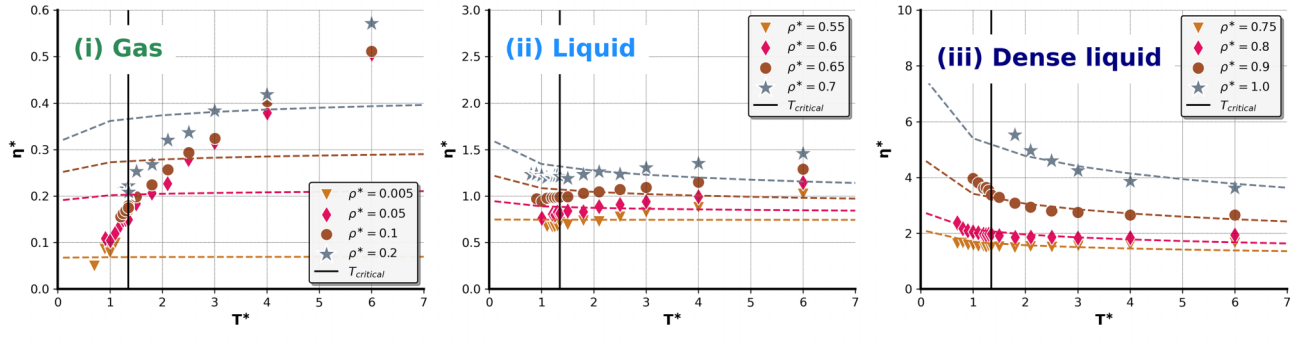
$$\eta_{1,\text{dilute}}^{*\text{symb}} = \lim_{\rho\to 0} \eta_1^{*\text{symb}} = a_1\sqrt{T^*}, \tag{26}$$

$$\eta_{2,dilute}^{*\text{symb}} = \lim_{\rho\to 0} \eta_2^{*\text{symb}} = 0, \tag{27}$$

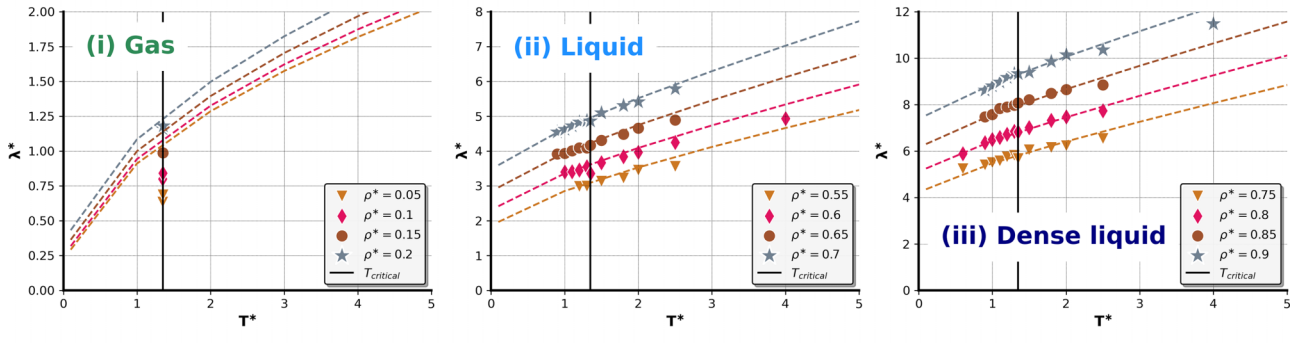
$$\lambda_{1,\text{dilute}}^{*\text{symb}} = \lim_{\rho\to 0} \lambda_1^{*\text{symb}} = c_1\sqrt{T^*}. \tag{28}$$



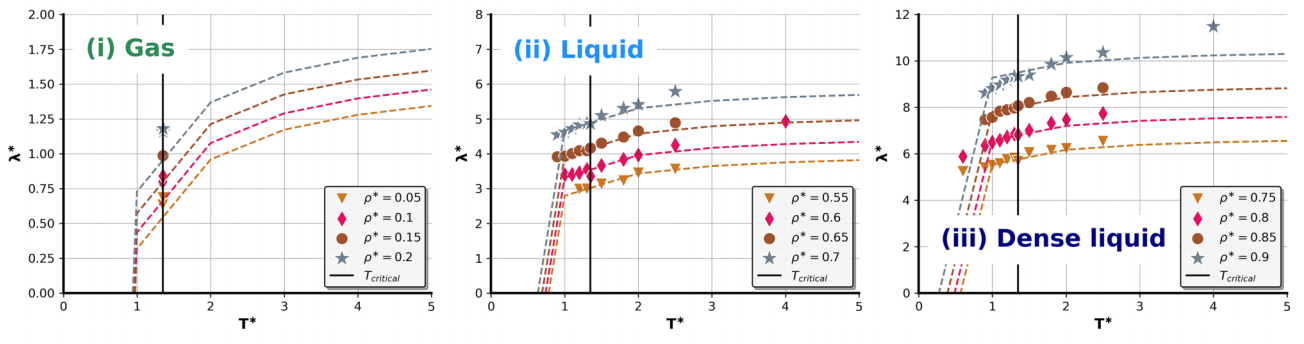
(a) Equation η_1^{*symb}



(b) Equation η_2^{*symb}



(c) Equation λ_1^{*symb}

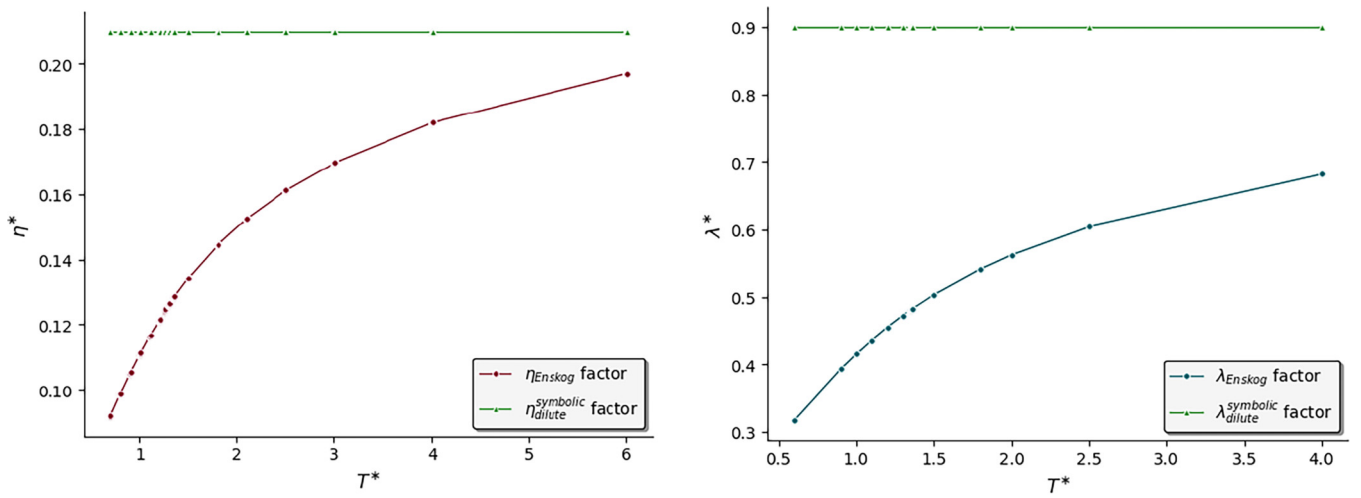


(d) Equation λ_2^{*symb}

FIG. 4. Fitting of symbolic equations at various fluid states. Color lines correspond to the respective equation data, black line is the T_c limit, and symbols to the available data points.

From the analysis above, it appears that $\eta_{2,dilute}^{*\text{symb}}$ equals to zero at dilute gas state, and this shows a nonphysical behavior. Thus, for viscosity estimation, only $\eta_1^{*\text{symb}}$ is kept next for further investigation.

The next step on establishing a firm physical basis to our derived expressions for viscosity and thermal conductivity, is their performance evaluation in various conditions and their relationship with some long-established theoretical models. We mostly focus on well-known approaches across various



(a) Viscosity (b) Thermal conductivity

FIG. 5. Performance of viscosity and thermal conductivity components (L, M, a_1, c_1) in the dilute situation.

fluid regions, without diving deep into a comprehensive presentation of those. More details can be found on the respective literature sources.

Starting from the dilute gas scenario, both viscosity ($\eta^{*\text{symb}}_{1,\text{dilute}}$) and thermal conductivity ($\lambda^{*\text{symb}}_{1,\text{dilute}}$) equations, suggest a square root dependence between the transport property and temperature. This term is also found in the Enskog theory for viscosity [Eq. (7)] and thermal conductivity [Eq. (15)].

Furthermore, one could apply slight modifications on the traditional theoretical approaches to gain more insight on the implied mechanisms. By reconfiguring the parts of Enskog expressions for viscosity and thermal conductivity [Eq. (7) and Eq. (15)], the first component of each equation would translate to the L and M terms, respectively, as

$$\eta_{\text{Enskog}} = \overbrace{\left(\frac{5}{16\sqrt{\pi}}\frac{1}{\Omega^{(2,2)*}}\right)}^{L}\left[\frac{\sqrt{mk_B}}{\sigma^2}\sqrt{T}\right], \tag{29}$$

$$\lambda_{\text{Enskog}} = \overbrace{\left(\frac{75}{64\sqrt{\pi}}\frac{1}{\Omega^{(2,2)*}}\right)}^{M}\left[\frac{k_B^{(3/2)}}{\sqrt{m}\sigma^2}\sqrt{T}\right]. \tag{30}$$

If we turn our attention to Eq. (26) for viscosity and Eq. (28) for thermal conductivity, substitute the reduced parameters with those of Eq. (6), and rearrange some parameters, we obtain

$$\eta^{\text{symbolic}}_{\text{dilute}} = (a_1)\left[\frac{\sqrt{mk_B}}{\sigma^2}\sqrt{T}\right], \tag{31}$$

$$\lambda^{\text{symbolic}}_{\text{dilute}} = (c_1)\left[\frac{k_B^{(3/2)}}{\sqrt{m}\sigma^2}\sqrt{T}\right]. \tag{32}$$

Notwithstanding the fact that the proposed SR-derived expressions capture exactly the dynamics of the second part of each equation, we have also calculated the values of L and M from Eqs. (29) and (30). For the estimation of $\Omega^{(2,2)*}$, we have utilized the commonly used empirical relation [62]

$$\Omega^{(l,s)*} = [A(T^*)^{-B}] + C[\exp(-DT^*)] + E[\exp(-FT^*)], \tag{33}$$

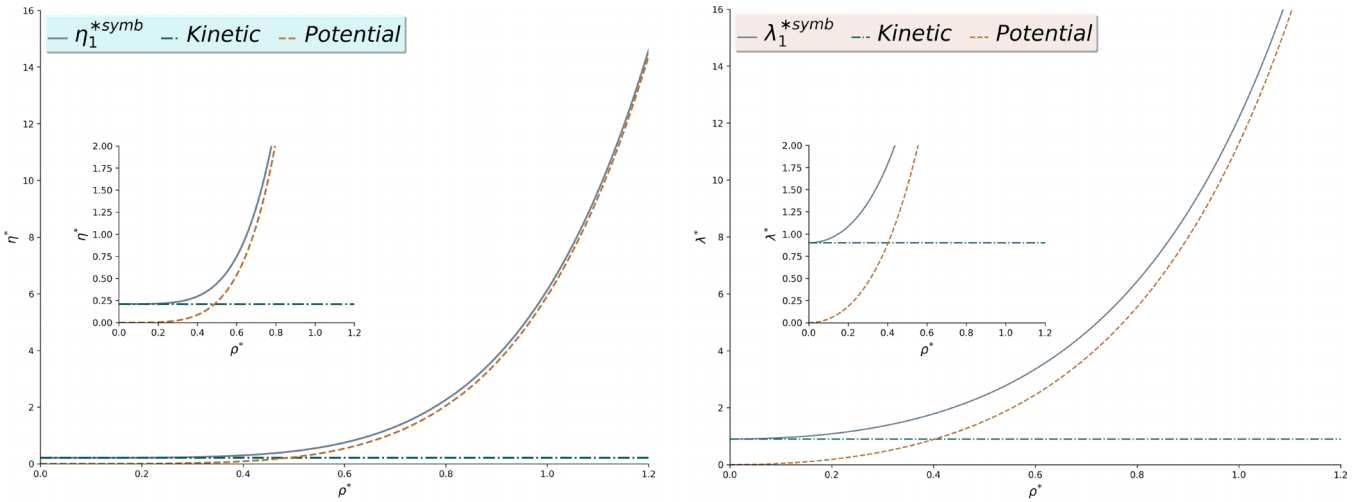
where, for the $l = s = 2$ situation, $A = 1.16145$, $B = 0.14874$, $C = 0.52487$, $D = 0.77320$, $E = 2.16178$, $F = 2.43787$, and T^* is given by Eq. (6). This equation is valid in the $0.3 \leqslant T^* \leqslant 100$ regime. Since our data points are inside this range, we have calculated $\Omega^{(2,2)*}$ at each data point for both transport properties. Considering that the $\Omega^{(2,2)*}$ factor of Enskog's expression is temperature dependant, a volatile response of the L and M parts is expected. On the contrary, symbolic expressions suggest a constant behavior of the corresponding factors (a_1 and c_1). Results appear in Fig. 5. Firstly, it is observed that both symbolic and Enskog factors operate in the same order of magnitude; however, deviations are evident for both instances. We can see that as temperature increases symbolic predictions converge towards Enskog's values while the differences are larger at lower temperatures. We can notice again that the differences are larger in the case of thermal conductivity where fewer data were available.

What is also worth mentioning, is that according to theoretical approaches [73,75], shear viscosity and thermal conductivity consist of two parts, a kinetic and a potential, with the kinetic part posing major influence over the transport property at gas phase, whereas at liquid phase the potential part conquers. The interesting aspect here is that $\eta_1^{*\text{symb}}$ and $\lambda_1^{*\text{symb}}$ somehow consists of two parts X_1, X_2 & Y_1, Y_2, respectively, as:

$$\eta_1^{*\text{symb}} = \begin{cases} X_1 = a_1\sqrt{T^*}, \\ X_2 = a_2\ \rho^{*4}\left(a_3 + \frac{a_4\,\rho^*}{\sqrt{T^*}}\right)^2. \end{cases} \tag{34}$$

$$\lambda_1^{*\text{symb}} = \begin{cases} Y_1 = c_1\sqrt{T^*}, \\ Y_2 = \rho^{*2}(T^* + \rho^* + e^{2\rho^*} + c_2). \end{cases} \tag{35}$$

Further analysis of $\eta_1^{*\text{symb}}$ and $\lambda_1^{*\text{symb}}$ by examining separately the behavior of $X_1 - X_2$ and $Y_1 - Y_2$ parts at different densities for constant temperature is made in Fig. 6. For



(a) Equation η_1^{*symb} (b) Equation λ_1^{*symb}

FIG. 6. Decomposing viscosity and thermal conductivity SR-extracted equations. Kinetic parts refer to X_1 and Y_1 and potential parts to X_2 and Y_2. Figures on the inset zoom into the points where transport properties increase.

simplicity, we consider $T^* = 1.0$. Since transport property value range may be misleading in the low density regime, an additional inset figure is presented inside the main plot.

Here, it is observed that for low densities, i.e., $\rho^* < 0.2$, (at gas phase), the overall behavior of $\eta_1^{*\text{symb}}$ resembles the behavior of X_1. For $0.2 < \rho^* < 0.5$, there is a transition phase where both parts contribute at the final value of viscosity, while at density values above 0.5 (at liquid phase), the behavior of viscosity is completely governed by X_2. This could be an indication that X_1 and X_2 operate as the kinetic and potential part respectively. To further support that assumption, the fact that X_1 has a strong relation with the Chapman-Enskog equation (i.e., the kinetic part) strengthens our hypothesis. In contrast to viscosity dynamics, $\lambda_1^{*\text{symb}}$ starts to approach Y_2 from a density value $\rho^* \approx 0.1$, i.e., when fluid is at the gas phase. Despite the Y_1 relation with the Chapman-Enskog equation for thermal conductivity, this characteristic complicates the behavior and requires further investigation. Therefore, for the next step, we plot the potential parts of generated expressions, i.e, X_2 and Y_2 at different fluid phases, similar to Fig. 4.

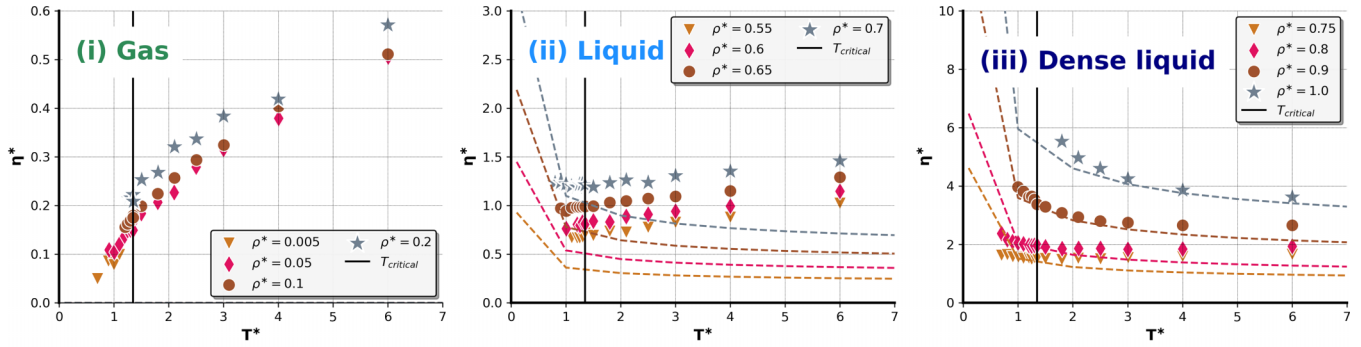
Furthermore, Fig. 7 demonstrates the performance of X_2 and Y_2 across different fluid regimes. The viscosity's ($\eta_1^{*\text{symb}}$) term, X_2, misses the fluid behavior at the gas phase and remains near to zero viscosity values. Since the potential part has a minor influence over the kinetic at the gas phase and the fact that X_1 resembles the kinetic, then X_2 correctly fails to approach gas measurements or, even more, its overall dynamics. Conversely, as the fluid becomes denser, X_2 achieves an increasingly fine fitting while concurrently sustaining physical correspondence, as it predicts viscosity reduction with temperature increase. Thus, it seems that the constitutive parts of viscosity, X_1 and X_2, exhibit strong similarity to the expected behavior of kinetic and potential parts.

In contrast, the Y_2 part of thermal conductivity ($\eta_1^{*\text{symb}}$) has shown less precision. It has to be noted that sparse dataset points at the gas phase, and the fact that they appear in the vicinity of the critical point, does not allow a direct estimation. However, what is evident is the tendency of Y_2 to increase at the gas phase. This fact contradicts to the expected behavior. Moreover, we notice that Y_2 perceives the overall physics at liquid and dense liquid phase, as it follows a trail similar to the simulation data points. Nonetheless, Y_2 performance seems to be significantly reduced, as it fails to capture the simulation data points. It should be reminisced that in Fig. 4 the united response of Y_1 and Y_2 achieves an excellent fit at both liquid and dense liquid phases. Therefore, Y_1, which resembles the Chapman-Enskog relation for dilute gas thermal conductivity, has a strong influence at the thermal conductivity in liquid and dense liquid phase. Therefore, it is deduced that $\lambda_1^{*\text{symb}}$ is not in fact two distinct parts, but these are interconnected and operate as one.

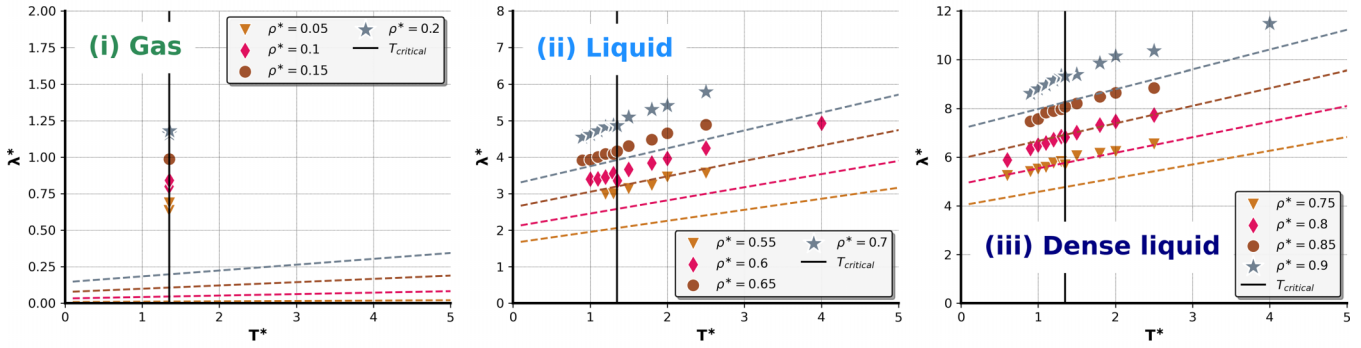
To sum up, from the initial pool of $\eta_1^{*\text{symb}}$, $\eta_2^{*\text{symb}}$, $\lambda_1^{*\text{symb}}$, $\lambda_2^{*\text{symb}}$ expressions, despite fine agreement with observation values, $\eta_2^{*\text{symb}}$ and $\lambda_2^{*\text{symb}}$ are discarded due to physical disagreement. On the other hand, $\eta_1^{*\text{symb}}$ has been found to achieve excellent physical and statistical agreement. By splitting its two constitutive parts into a density independent kinetic part, X_1, and a temperature dependent potential part, X_2, it has been shown that these two have diametrically opposite characteristics and strongly resemble the kinetic and potential part that the DFT approach suggests. In addition, resemblance of the X_1 term to the Chapman-Enskog relation for dilute viscosity [Eq. (7)], strengthens this hypothesis.

On the contrary, this scenario does not apply to $\lambda_1^{*\text{symb}}$, in which the two constitutive terms, Y_1 and Y_2, operate as one. However, since the thermal conductivity dataset contains only a small set of observations at the gas phase, near the critical point, $\lambda_1^{*\text{symb}}$ would need more data points to be established as a firm relation for estimating thermal conductivity of fluids. At present, it can be considered as a decent choice for the liquid and dense liquid states.

It is important to emphasize, that the derived equations are solely derived from macroscopic fluid properties (density



(a) Part X_2 of equation η_1^{*symb}



(b) Part Y_2 of equation λ_1^{*symb}

FIG. 7. Parts X_2 and Y_2 of viscosity and thermal conductivity equations ($\eta_1^{*\text{symb}}$ and $\lambda_1^{*\text{symb}}$) at different fluid phases.

and temperature). Comparison to GK formulas, Eq. (9) and Eq. (19), is not possible, since these equations contain terms from the microscopic state of the system, such as particle positions, velocities, potential energies, and their correlated time averages. The proposed SR-based equations rely on macroscale properties and pose as a fast and accurate alternative means of transport properties calculation, making them practical tools for various applications. Nevertheless, more GK or relevant methods simulation data would train the SR model even better.

Therefore, from a physics point of view, the proposed relations reflect the underlying physical mechanisms that affect the transport properties of fluids, with a closed form equation that seems robust and accurate. Efforts to obtain closed form relations from basic assumptions have been made, as for example, the Enskog relations, but these are based on simple assumptions, while the real situation may be more complex, and interdependencies cannot be easily incorporated. With aid of SR, such a closed form relation is able to catch the physical meaning in a transparent way, which is beneficial compared to other ML methods, such as shallow and deep learning. Comparison with semianalytical equations both shows the accuracy of the approximations and extends the description to more complex situations, which cannot be treated with other approaches.

V. CONCLUSION

Machine learning techniques have entered almost every aspect of computational science research, providing enhanced estimation with increased accuracy and minimal computational cost. A symbolic regression model has been chosen here, capable of unravelling hidden relations inside a simulation dataset and providing physical interpretation, overpassing common black-box models. Viscosity and thermal conductivity equations are extracted, without previous knowledge of the system under investigation. Results indicate that it is possible to obtain macroscopic equations driven by microscopic state data and achieve satisfying accuracy, while concurrently adhering to statistical and physical laws.

The proposed viscosity equation, $\eta_1^{*\text{symb}}$, presents an interesting behavior bound to physical intuition, with one term representing the gas phase behavior where viscosity increases with temperature, and a second term which encloses a combined effect of density and temperature. In the liquid phase, a phenomenon reproduced by our equation lies upon the increased value of viscosity, and at the same time, a viscosity reduction with an increase of temperature at constant density. Another interesting aspect is that the gas related term ($\sqrt{T}$) is similar to the Enskog relation. As far as thermal conductivity is concerned, the proposed equation $\lambda_1^{*\text{symb}}$ has a term with increasing temperature behavior (depending only on temperature), along with combined density/temperature terms, and a term based only on density.

Going one step further from conventional models, the generated expressions encompass microscopic molecular dynamics accuracy with general macroscopic features, and provide a reliable model able to bridge dynamics across scales. Viscosity expressions adhere to the overall behavior of traditional empirical and theoretical models. On the other

hand, due to limited data, even though thermal conductivity expressions achieve fine accuracy and low error measures, it is harder to establish a physical interpretation to existing domain knowledge.

Finding the golden mean between computationally intensive simulations and ML-assisted schemes could be the dominant approach in computational science in the near future.

[1] S. Bastea, Phys. Rev. E **68**, 031204 (2003).
[2] E. Ermakova, J. Solca, H. Huber, and M. Welker, J. Chem. Phys. **102**, 4942 (1995).
[3] H. Zhang, B. jian Zhang, S. Liang, Y. Lu, W. Hu, and Z. Jin, Chem. Phys. Lett. **350**, 247 (2001).
[4] D. M. Heyes and A. C. Brańka, Phys. Chem. Chem. Phys. **10**, 4036 (2008).
[5] H. Stassen and W. A. Steele, J. Chem. Phys. **102**, 932 (1995).
[6] M. Frank and D. Drikakis, Entropy **20**, 362 (2018).
[7] H. Zhang, H. Ye, Y. Zheng, and Z. Zhang, Microfluid. Nanofluid. **10**, 403 (2011).
[8] S. H. Jamali, R. Hartkamp, C. Bardas, J. Sohl, T. J. Vlugt, and O. A. Moultos, J. Chem. Theory Comput. **14**, 5959 (2018).
[9] G. Galliero and C. Boned, Phys. Rev. E **79**, 021201 (2009).
[10] F. Sofos, T. Karakasidis, and A. Liakopoulos, Int. J. Heat Mass Transfer **52**, 735 (2009).
[11] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction and Diffusion in Gases* (Cambridge University Press, Cambridge, 1990).
[12] Y. Rosenfeld, Phys. Rev. A **15**, 2545 (1977).
[13] J. C. Dyre, J. Chem. Phys. **149**, 210901 (2018).
[14] I. H. Bell, R. Messerly, M. Thol, L. Costigliola, and J. C. Dyre, J. Phys. Chem. B **123**, 6345 (2019).
[15] B. Song, X. Wang, and Z. Liu, Mol. Simul. **42**, 9 (2016).
[16] J. A. Barker, W. Fock, and F. Smith, Phys. Fluids **7**, 897 (1964).
[17] A. Aubreton and M.-F. Elchinger, J. Phys. D **36**, 1798 (2003).
[18] B. Haghighi, M. Fathabadi, and M. Papari, Fluid Phase Equilib. **203**, 205 (2002).
[19] T. Hosseinnejad, H. Behnejad, and V. H. Shahmir, Fluid Phase Equilib. **258**, 155 (2007).
[20] D. Mohammad-Aghaie, E. Sookhaki, and F. Zargari, Phys. Chem. Res. **7**, 271 (2019).
[21] E. L. Shock and H. C. Helgeson, Geochim. Cosmochim. Acta **52**, 2009 (1988).
[22] M. Anouti, J. Jacquemin, and P. Porion, J. Phys. Chem. B **116**, 4228 (2012).
[23] B. Todd and J. Young, J. Power Sources **110**, 186 (2002).
[24] A. Sohn and C. Yu, Mater. Today Phys. **19**, 100433 (2021).
[25] Y. Zhu, B. Wei, J. Liu, N. Z. Koocher, Y. Li, L. Hu, W. He, G. Deng, W. Xu, X. Wang, *et al.*, Mater. Today Phys. **19**, 100428 (2021).
[26] T. H. Chung, M. Ajlan, L. L. Lee, and K. E. Starling, Ind. Eng. Chem. Res. **27**, 671 (1988).
[27] J. F. Ely and H. Hanley, Ind. Eng. Chem. Fundam. **20**, 323 (1981).
[28] G. M. Geise, D. R. Paul, and B. D. Freeman, Prog. Polym. Sci. **39**, 1 (2014).
[29] P. G. Radaelli, G. Iannone, M. Marezio, H. Y. Hwang, S.-W. Cheong, J. D. Jorgensen, and D. N. Argyriou, Phys. Rev. B **56**, 8265 (1997).
[30] J. P. Meyer, S. A. Adio, M. Sharifpur, and P. N. Nwosu, Heat Transfer Eng. **37**, 387 (2016).
[31] L. S. Sundar, K. Sharma, M. Naik, and M. K. Singh, Renewable Sustainable Energy Rev. **25**, 670 (2013).
[32] M. Tham and K. Gubbins, Ind. Eng. Chem. Fundam. **9**, 63 (1970).
[33] Y. Takeshita, K. Shimamura, S. Fukushima, A. Koura, and F. Shimojo, J. Phys. Chem. Solids **163**, 110580 (2022).
[34] P. L. Silvestrelli, J. Chem. Phys. **158**, 134503 (2023).
[35] Z. Sun, Y. Kang, and S. Li, Chem. Eng. Sci. **277**, 118847 (2023).
[36] P. Rehner, G. Bauer, and J. Gross, Ind. Eng. Chem. Res. **62**, 5347 (2023).
[37] M. Frank, D. Drikakis, and V. Charissis, Computation **8**, 15 (2020).
[38] M. Schmid, D. Altmann, and G. Steinbichler, Polymers **13**, 2652 (2021).
[39] G. D. Goh, S. L. Sing, and W. Y. Yeong, Artif. Intell. Rev. **54**, 63 (2021).
[40] D. Morgan and R. Jacobs, Annu. Rev. Mater. Res. **50**, 71 (2020).
[41] D. Reimann, K. Nidadavolu, H. ul Hassan, N. Vajragupta, T. Glasmachers, P. Junker, and A. Hartmaier, Front. Mater. **6**, 181 (2019).
[42] F. Sofos, C. Stavrogiannis, K. K. Exarchou-Kouveli, D. Akabua, G. Charilas, and T. E. Karakasidis, Fluids **7**, 116 (2022)
[43] D. Angelis, F. Sofos, and T. E. Karakasidis, Arch. Comput. Methods Eng. **30**, 3845 (2023).
[44] E. Kabliman, A. H. Kolody, M. Kommenda, and G. Kronberger, AIP Conf. Proc. **2113**, 180009 (2019).
[45] W. Gilpin, Chaos as an interpretable benchmark for forecasting and data-driven modelling, in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, edited by J. Vanschoren and S. Yeung (Curran, 2021), Vol. 1.
[46] T. Kitano, T. Kataoka, and T. Shirota, Rheol. Acta **20**, 207 (1981).
[47] S. Bair and P. Kottke, Tribol. Trans. **46**, 289 (2003).
[48] J. Ascough, R. G. Chapman, and N. H. March, Phys. Chem. Liq. **18**, 253 (1988).
[49] K. Rah and B. C. Eu, Phys. Rev. E **60**, 4105 (1999).
[50] F. Yoshida, Phys. Lett. A **117**, 77 (1986).
[51] B. A. Younglove and H. J. M. Hanley, J. Phys. Chem. Ref. Data **15**, 1323 (1986).
[52] H. J. M. Hanley, R. D. McCarty, and W. M. Haynes, J. Phys. Chem. Ref. Data **3**, 979 (1974).
[53] N. Ohtori and Y. Ishii, J. Chem. Phys. **143**, 164514 (2015).
[54] K. Papastamatiou, F. Sofos, and T. E. Karakasidis, AIP Adv. **12**, 025004 (2022).
[55] F. Sofos, A. Charakopoulos, K. Papastamatiou, and T. E. Karakasidis, Phys. Fluids **34**, 062004 (2022).
[56] T. M. Alam, J. P. Allers, C. J. Leverant, and J. A. Harvey, J. Chem. Phys. **157**, 014503 (2022).
[57] T. E. Karakasidis, F. Sofos, and C. Tsonos, Fluids **7**, 321 (2022).

[58] C. Loftis, K. Yuan, Y. Zhao, M. Hu, and J. Hu, J. Phys. Chem. A **125**, 435 (2021).
[59] C. D. Muzny, M. L. Huber, and A. F. Kazakov, J. Chem. Eng. Data **58**, 969 (2013).
[60] M. Izadmehr, R. Shams, and M. H. Ghazanfari, J. Nat. Gas Sci. Eng. **30**, 364 (2016).
[61] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954).
[62] B. E. Poling, J. M. Prausnitz, and J. P. O'connell, *Properties of Gases and Liquids* (McGraw-Hill Education, New York, 2001).
[63] I. W. Smith, *Kinetics and Dynamics of Elementary Gas Reactions: Butterworths Monographs in Chemistry and Chemical Engineering* (Butterworth-Heinemann, London, Boston, 1980).
[64] G. Rutkai, M. Thol, R. Span, and J. Vrabec, Mol. Phys. **115**, 1104 (2017).
[65] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, 2nd ed. (Oxford University Press, Oxford, 2017).
[66] K. Meier, A. Laesecke, and S. Kabelac, J. Chem. Phys. **121**, 3671 (2004).
[67] M. Bugel and G. Galliero, Chem. Phys. **352**, 249 (2008).
[68] D. S. Viswanath, T. K. Ghosh, D. H. Prasad, N. V. Dutt, and K. Y. Rani, *Viscosity of Liquids: Theory, Estimation, Experiment, and Data* (Springer Science & Business Media, Dordrecht, 2007).
[69] G. Fernández, J. Vrabec, and H. Hasse, Fluid Phase Equilib. **221**, 157 (2004).
[70] B. D. Todd, D. J. Evans, and P. J. Daivis, Phys. Rev. E **52**, 1627 (1995).
[71] K. P. Travis, B. D. Todd, and D. J. Evans, Phys. Rev. E **55**, 4288 (1997).
[72] F. Sofos, T. Karakasidis, and A. Liakopoulos, Int. J. Heat Mass Transfer **53**, 3839 (2010).
[73] A. E. Nasrabad, R. Laghaei, and B. C. Eu, J. Phys. Chem. B **109**, 8171 (2005).
[74] R. Laghaei, A. E. Nasrabad, and B. C. Eu, J. Phys. Chem. B **109**, 5873 (2005).
[75] K. Rah and B. Chan Eu, J. Chem. Phys. **115**, 9370 (2001).
[76] S. Stephan, M. Thol, J. Vrabec, and H. Hasse, J. Chem. Inf. Model. **59**, 4248 (2019).
[77] G. C. McNeil-Watson and N. B. Wilding, J. Chem. Phys. **124**, 064504 (2006).
[78] T. Tohme, D. Liu, and K. Youcef-Toumi, arXiv:2205.15569.
[79] T. Back, *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms* (Oxford University Press, New York, 1996).
[80] J. R. Koza, Stat. Comput. **4**, 87 (1994).
[81] C.-Y. Liu and T. P. Senftle, Curr. Opin. Chem. Eng. **37**, 100832 (2022).
[82] M. Cranmer, MilesCranmer/PySR v0.2, Zenodo (2020), doi: 10.5281/zenodo.4041459.
[83] C. Fox, N. Tran, N. Nacion, S. Sharlin, and T. R. Josephson, arXiv:2301.11919.