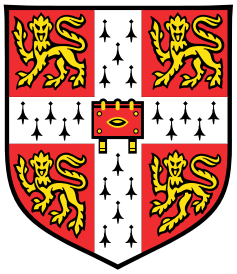
Physical properties of alkanes and their mixtures



Pavao Santak

Department of Physics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Fitzwilliam College September 2021

To my family.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 60,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

This Thesis contains the material published in the peer-reviewed literature

- **Chapter 2:** P. Santak and G. Conduit. *Predicting physical properties of alkanes with neural networks.* Fluid Phase Equilibria, **501** (112259), 2019.
- **Chapter 3** P. Santak and G. Conduit. *Enhancing NEMD with automatic shear rate sampling to model viscosity and correction of systematic errors in modeling density: Application to linear and light branched alkanes.* The Journal of Chemical Physics, **153(1),** 2020.

Pavao Santak
September 2021

Acknowledgements

Despite its reputation as an apprenticeship in academic research, a PhD is as much a personal journey as it is an academic one. With this in mind, there are many people to thank, as without their help and support, a PhD would be a much more gruelling experience.

I would first like to thank my supervisor, Gareth Conduit for encouraging me participate in a diverse set of activities to gain additional transferable skills, teaching me the importance of getting things done in a timely fashion, and the significance of clearly explaining scientific concepts to non-experts. Additionally, I would like to thank the BP technologists, Les Bolton, Nikos Diamantonis, Corneliu Buda, Phil Davies, and John Baker, for providing an amazing technical and administrative support during the course of this project.

Next, I want to thank the Master, the Fellows, and Members of Fitzwilliam College for providing the best living experience possible. Special thanks go to some of the people I have had a privilege to collaborate with while I was on the Fitzwilliam MCR Committee. To Andrew Powell, for teaching me the importance of constructive dialogue and consummate professionalism, To André, for showing me the importance of tact and diplomacy, and Ellie, for her transparency and honesty, showing me that they're worth fighting for irrespective of the odds.

I have to thank many of my friends, who have made my time in Cambridge better than I could have ever imagined, and provided me with personal lessons I will never forget. To TMB, for being an amazing friend and housemate since Day 1 on more occasions than I can count. To Nik, Jamie, and Oliver, for all the fun times in and out of Cambridge and the hope we stay in touch even if we end up living on different continents. To Tom and Pete, for taking the already unique life in 114 Huntingdon Road to another level with the dinner parties and unconventional dressing styles. To Conor, for being a great self-appointed PR manager, co-inventing a new sport, and providing an endless source of laughter. To Dom and Dan, for being amazing, supportive, and non-judgemental friends, as well as taking part in many conversations I cannot tell you about in this Thesis. To Numerate Derek, for his unique open mind and hopefully many more years of exciting friendship. To Petr, for being a good, reliable friend, and a good squash opponent. To Owen, for amazing times and lovely garden conversations during the first lockdown, Sage, for showing me that true bohemians

can be found even in a lake of standard people, and Jon, for his lovely presence and making the second lockdown much easier. To Shelby, for being a constant source of optimism, the unforgettable deep meaningful conversations and bike rides. To Aisha, our Tom, and Victoria, for their tolerance and patience, as well as a glimpse of normal life, even if I'm not sure I like it.

I would also like to thank the people who have left a lasting impression on me and gave me the motivation for future endeavours, but whom I have unfortunately not had a chance to spend enough time with. To David, for showing me the power of healthy obsessions and taking an outside view in all matters of life, Maya and Gabby, for teaching me the importance of an open mind and unbiasedness, and Derek (British), for giving the terms collaboration and integrity new, modern meanings.

Finally, I would like to thank my family for unparalleled support since the beginning of my studies and enduring my almost continued absence during the last several years, and to my Croatian friends for keeping in touch despite the fact that our lives are by now so different that it occasionally seems like we live in a different universe.

Abstract

Alkanes and their mixtures are some of the physically simplest molecules and are widely used in industry, yet the connection between their structure and physical properties is still poorly understood. To make progress, we study the properties of pure alkanes with neural networks and molecular dynamics, while we develop a new theoretical framework to study the properties of mixtures of alkanes.

We first encode alkanes' structure into five non-negative integers and use them as neural network inputs. Then, we utilize the neural networks to study the boiling point, vapor pressure, heat capacity, and melting point of light alkanes, as well as flash point and kinematic viscosity of linear alkanes. Neural networks model all these properties more accurately than the competing statistical and physico-chemical methods, while the cross-validation results indicate that they can confidently and accurately extrapolate the boiling point, heat capacity, and vapor pressure models to heavy alkanes. Still, due to a lack of experimental data for non-linear alkanes, neural network flash point and kinematic viscosity models cannot extrapolate to heavy alkanes, while the comparatively low accuracy of melting point models relative to other properties' models suggests that additional physical effects need to be incorporated into them.

To obtain synthetic data as a supplement for the experimental kinematic viscosity dataset, we perform molecular dynamics simulations for density and non-equilibrium molecular dynamics (NEMD) simulations for dynamic viscosity. Density simulation results are corrected through a data-driven approach to increase their accuracy, and we develop a sampling algorithm that automatically selects the shear rates at which to perform the viscosity simulations. The sampling algorithm is tested on linear alkanes, and simulation results are in excellent agreement with the experiments, encouraging applications to more complex alkanes.

Then, we use neural networks with molecular structure as inputs to model the molecular dynamics density simulation values and extrapolate to 11-heptyltricosane, 8,11-dipentyloctadecane, and 8,14-dipentylhenicosane at 40°C and 100°C. These extrapolated density values are used as state points for the NEMD viscosity simulations, which are performed with the help of the shear rate sampling algorithm. While the accuracy of neural network models is high, and the usefulness and reliability of the sampling algorithm is further established, viscosity

simulation results are not in a good agreement with the experiment due to systematic error in the force field.

Finally, to model properties of mixtures of alkanes, we develop a theory of mixtures whose molecules' positions have a uniform spatial distribution. We apply this theory to molar volume, isentropic compressibility, surface tension, and dynamic viscosity of mixtures of alkanes, first by fitting to experimental data, and then by using the best fit parameters for viscosity to predict viscosity of further mixtures. Best fits and predictions show excellent agreement with the experiments, and our theory shows promise for further applications to mixtures of alkanes, while its conceptual basis has the potential to be applied to other types of mixtures as well.

Table of contents

List of figures

List of tables

Chapter 1

Background

1.1 Introduction

Physical sciences encompass the most successful set of methods to study the material non-living natural phenomena. These methods have enabled humanity to better understand the universe around us, from the microscopic atoms and molecules up to macroscopic planets and galaxies. Beyond their epistemic value, advances in physical sciences have led to innovations that have altered the course of the human civilization. From the personal computers to automobiles, modern society would be unimaginable without inventions stemming from the physical sciences.

One of the most important applications of physical sciences are the discovery and production of materials, which has enabled the development of essential modern machines and devices, such as airplanes or printers. With a continued increase in world population and the per capita consumption, the search for improved materials will underpin the attempts to create a more resource efficient society. Among the many materials that will undergo a major evolution in the coming decades are the industrial lubricants. They are already an important component of the modern industry, helping protect the surfaces that are in contact with each other, transfer heat, and reduce corrosion and friction. Even as the world moves away from the carbon based energy products, the demand for them is expected to grow [1].

1.2 Alkanes

A demand for improved lubricants is accompanied by the need for deeper understanding of their physics. This necessitates a deeper investigation of their components' properties. It

[1] https://www.grandviewresearch.com/industry-analysis/lubricants-market

is known that the lubricants are produced by combining a base oil with additives to obtain the desired properties. The majority of base oils comprise alkanes [82][7] (Figure 1.1), the are hydrocarbons in which the carbon atoms are connected by a single covalent bond [21][115](Figure 1.1).

Alkanes are mostly found as components of crude oils, which are organic materials usually found in the Earth's crust, and made by disintegration of the fossilised organic materials under high pressure and temperatures without the presence of oxygen. Consequently, alkanes have a range of applications in the oil and gas industry. The lightest alkanes, methane, ethane, propane, and butane are used for heating and cooking purposes, and alkanes with between five and eight carbon and with between nine and sixteen carbon atoms are essential elements of petrol and jet fuels. Beyond their immediate applications, they are often used as a basis for production of various organic compounds [17], while there has been a growing interest in their role in the chemistry of atmosphere [108].

Alkanes are commonly split into three categories, which differ by the number of carbon atoms connected to each other:

- Linear alkanes, where two carbon atoms are connected to one, and the rest to two carbon atoms.
- Branched alkanes, where two carbon atoms are connected to one, and the rest to two or three carbon atoms.
- Cycloalkanes, where each carbon atom is connected to two carbon atoms.

One can study either alkanes properties when they react with other components (chemical properties), or independent of them (physical properties). In this Thesis, we study their physical properties of linear and branched alkanes with fewer than three branches. Specifically, we investigate the properties that are important for assessing the performance of base oil lubricants: melting point, flash point, boiling point, vapour pressure, density, and dynamic viscosity.

1.3 Artificial neural networks

1.3.1 Linear regression

Physical properties of most alkanes have not been measured, since alkanes with more than twelve carbon atoms are rarely found as pure components in nature, forming mixtures with the other molecules instead. Additionally, while statistical mechanics [103] in principle offers

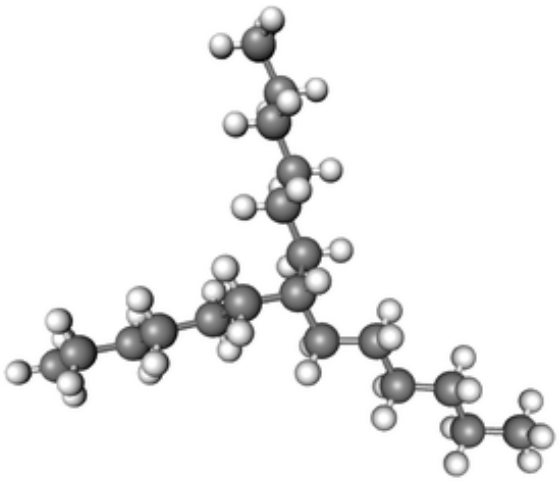


Fig. 1.1 An example of a branched alkane (7-hexyl-tridecane).

a suitable framework for theoretical studies of alkanes' properties, it has not yet produced a formalism capable of accurately describing them.

To make progress, we look for patterns in the existing alkanes' data and utilize computational statistics, which can learn and generalize from data to predict the unknown property values. To accomplish this, we propose a model $M(W)$ expecting that it accurately describes the mathematical relationship between the physically motivated independent variables ($\vec{x}$) and a physical property (y). Then, we analytically or numerically find the weights that best (W_{OPT}) describe this relationship by minimizing a measure of the distance ($\mu(W)$) between the experimental physical property values and their model approximations. To make predictions, the model with optimal weights is applied to independent variables with the unknown property values.

The simplest possible such model is linear. When the distance function is equal to half the average square difference between the properties and their model estimates, which is known as a mean square error (MSE) function, this formalism is called linear regression [57]

$$\begin{aligned} M(W) &= XW \\ MSE(W) &= \frac{1}{2N}(XW-Y)^T(XW-Y), \end{aligned} \tag{1.1}$$

where $M(W)$ is a N-dimensional column vector of property fits, X a $N \times (p+1)$ matrix comprising independent variable data entries augmented by a N-dimensional identity column vector to allow the modelled physical property values to be non-zero even when the values of

all the independent properties are zero, W a $(p+1)$-dimensional column vector of weights, Y a N-dimensional column vector of experimental property values, N a number of data entries, p a number of independent variables, and T the transpose of a matrix.

In linear regression, the optimal weights and their uncertainties can be found analytically

$$\begin{aligned} \mathbb{E}(W_{\mathrm{OPT}}) &= (X^T X)^{-1} X^T Y \\ \delta(W_{\mathrm{OPT}}) &= \left(\frac{1}{N-(p+1)} (X^T X)^{-1} I_{(p+1)\times 1} \sum_{i=1}^{N} (y_i - \hat{y})^2 \right)^{0.5^*}, \end{aligned} \tag{1.2}$$

where $I_{(p+1)\times 1}$ symbolises a $(p+1)$-dimensional identity column vector, i a data entry index, $\hat{y}$ an average of all the physical property values, and 0.5^* a square root function acting on a matrix element-wise.

1.3.2 Universal approximation property

Mathematical simplicity of linear regression relative to other modern statistical methods enables simpler analysis of its mathematical properties and the results' accuracy. However, linear regression cannot capture non-linear functional dependencies between the independent and dependent variables. Accordingly, unless we are able to identify a set of independent variables that are linearly correlated with a property of interest, linear regression cannot successfully model physical properties of alkanes.

Fortunately, we can build upon the idea of a linear model to capture the non-linear dependencies among the variables. To do so, we represent a linear model as a network. While this representation looks cumbersome, if we set

$$\tilde{X}^2 = \sigma(X^1 W^1), \tag{1.3}$$

where σ is a non-polynomial continuous function called the transfer function, acting on a matrix element-wise, X^1 is a $N \times (p+1)$ matrix of independent variables augmented with a N-dimensional column of ones and W^1 a matrix of hidden node weights, we construct an artificial neural network [57] [43] [14] (Figure 1.2). Note that X^2 is obtained by augmenting $\tilde{X}^2$ with an identity column vector to allow the modelled physical property values to be non-zero even when the values of all the independent properties are.

Non-linear transfer functions equip artificial neural networks with the theoretical ability to approximate the optimal functional relationship between the independent variables and the dependent variable to an arbitrary accuracy [31] [106] [52] [89], which is known as the universal approximation property (UAP). Coupled with the increasing power of computational

hardware, the UAP has helped neural networks become a cornerstone of the digital age and find applications in diverse scientific fields, from image recognition and computer vision [43] [123] to natural language processing [41].

Due to their versatility, neural networks can have many structures, which mutually differ by the number of hidden layers, types of connections and mathematical relationships between the nodes, and the transfer function(s) used. In this Thesis, we use a feedforward neural network with one hidden node (Figure 1.2) and the rectified linear unit transfer function

$$\mathrm{ReLu(x)} = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0. \end{cases} \tag{1.4}$$

.

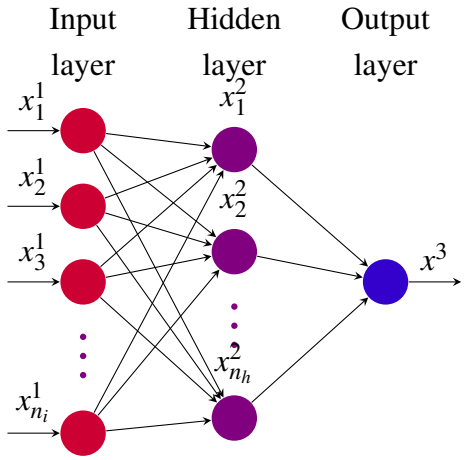


Fig. 1.2 Schematic of a feedforward neural network with one dependent variable. Circles represent networks' nodes, which themselves represent the variables. Variables are denoted by x_l^k, where x is the variable, k the layer index, l the node index, and x^3 the prediction for y.

1.3.3 Calculating the optimal weights

Approximating arbitrary functional relationships between variables is mathematically challenging. This is because the neural network MSE function (Equation 1.1) is non-convex, with multiple local minima, so a numerical scheme needs to be used to find the best weights. A numerical scheme is not guaranteed to find the global minimum. In practice, however, the values of the distance function around most local minima are close to the distance function value at a global minimum, so finding a "good" local minimum is sufficient [24].

Probably the most famous numerical algorithm to find the best weights, i.e. to train the network is the gradient descent [30], although the conjugate gradient [49] and the L-BFGS

[72] algorithms are frequently used as well. Before performing the gradient descent, we normalize all the variables through

$$\begin{aligned} \tilde{X}^1 &= \frac{\tilde{X}^1 - \mu_{\tilde{X}^1}}{\sigma_{\tilde{X}^1}} \\ Y &= \frac{Y - \mu_Y}{\sigma_Y}, \end{aligned} \tag{1.5}$$

where $\tilde{X}^1$ is the X^1 matrix without the last column, Y a set of dependent variable values, $(\mu_{\tilde{X}^1}, \sigma_{\tilde{X}^1})$ are matrices whose columns comprise the means and standard deviations of independent variables, and (μ_Y, σ_Y) are the mean and the standard deviation of the dependent variable. The normalization ensures all the independent variables are treated with equal importance, which helps us to find better weights more quickly.

We then initialize each weight with Gaussian random numbers with 0 mean and $\frac{2}{q}$ variance, where q is the number of hidden nodes. This weight initialization has been shown to be mathematically optimal for the ReLu function [45]. The weights are iteratively updated with a negative product of a positive real number α, known as a learning rate, and the gradient of the cost function with respect to the weights

$$W_{k+1} = W_k - \alpha \nabla_{W_k} \mu(W_k), \tag{1.6}$$

where k denotes the iteration step and ∇ the gradient. For the datasets used in this Thesis, we found the learning rate of 0.01 to be large enough to allow the gradient descent to find a good local minimum in a few enough steps but small enough not to "overshoot" good local minima and get stuck in a bad one. To find the gradient, we use a backpropagation algorithm [111], in which the partial derivatives of the MSE with respect to the weights are calculated through chain rule.

Unfortunately, since each weight needs to be updated separately, gradient descent is much slower than the L-BFGS or conjugate gradient algorithms. Nonetheless, we can speed it up, typically by $\mathcal{O}(10^2)$ through vectorization, which allows us to update all the weights at once. With vectorization, we can typically train a neural network in under a minute on datasets used in this Thesis. The vectorized form of gradient descent is

$$\begin{aligned} W^2_{k+1} &= W^2_k - \frac{\alpha}{N}(X^2_k)^T(X^3 - Y) \\ W^1_{k+1} &= W^1_k - \frac{\alpha}{N}(X^1)^T\left((\tilde{X}^2_k \tilde{W}^2_k - Y)(\tilde{W}^2_k)^T \odot \mathrm{H}(X^1 W^1_k)\right), \end{aligned} \tag{1.7}$$

where k denotes the iteration step , W^2 a set of weights in the output layer, W^1 a set of weights in the hidden layer, $\tilde{W}^2$ is W^2 without the final row, $\tilde{X}^2$ is given by Equation 1.3, X^2 is $\tilde{X}^2$ augmented by an identity column vector, $\odot$ the Hadamard product, and H the Heavyside function acting on a matrix element-wise.

The weights are updated until a convergence criterion is met or the maximum number of steps is performed. In this Thesis, we use the early stopping convergence criterion. In early stopping, we first split the data into k subsets, merge $k-1$ of them into a training set, and test the network's predictive power on the remaining set. We record the MSE's training and test set values every 250 steps. Once more than 5000 steps have been performed, we also start recording the averages of the last 20 training and test set errors. If this average has been decreasing for last 20 training recordings yet increasing for last 20 test recordings, we stop training the network. Otherwise, we train it for 200000 steps (Figure 1.3).

When we perform the above procedure k times, with each of the k subsets serving as a test set exactly once, we perform a k-fold cross-validation, which allows us to estimate the true predictive power of neural networks. In this Thesis, we usually use $k = 5$. With cross-validation, we can also find the optimal number of hidden nodes by fixing the data split and performing the cross-validation as a function of number of nodes. The optimal neural network structure is the one with the best cross-validation performance (Figure 1.3).

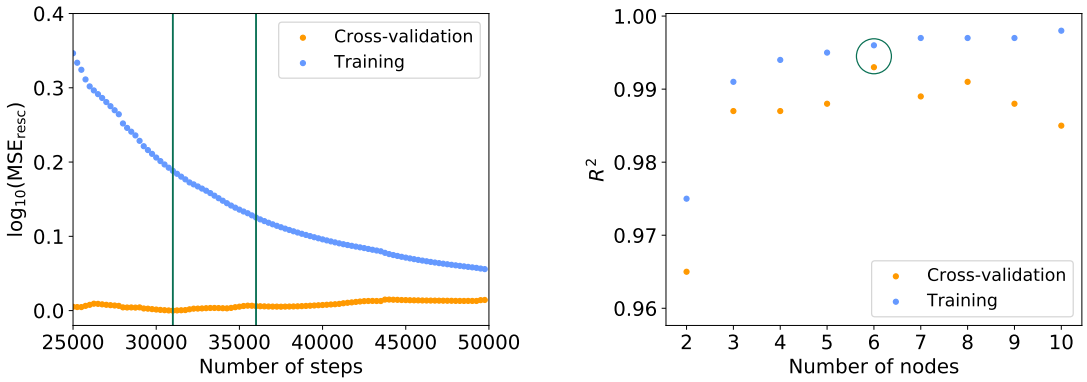


Fig. 1.3 Optimization of neural network's performance, illustrated on the boiling point of alkanes. Early stopping is illustrated on the left and the identification of optimal number of hidden nodes through cross-validation is illustrated on the right. For early stopping, only the 25000^{th} to 50000^{th} steps are shown. The logarithm under a base 10 of training and test MSE functions rescaled by their minima are presented. Test error is the smallest after 31000 steps so the would have stopped at 36000 steps had the early stopping been applied. To select the best number of nodes, we use a coefficient of determination R^2 as a metric (chapter 2), whose maximization is mathematically equivalent to the minimization of the MSE.

Even after normalizing the data, optimally initializing the weights, choosing a good learning rate, vectorizing the gradient descent, applying early stopping and selecting the optimal number of hidden nodes, we are not guaranteed to find a good local minimum, since an unlucky weight initialization can lead to gradient descent being stuck in a saddle point region. To resolve this issue, we train 16 neural networks in parallel, each with a different fixed random seed, and take results from only 6 of them. A different constant random seed for each network is chosen, which ensures reproducibility of results. With a multiple good set of weights, we can reliably calculate the expected value and the uncertainty in neural network fits/predictions.

1.4 Molecular dynamics

1.4.1 Motivation

Unfortunately, neural network still cannot predict all the properties of large alkanes, as insufficient experimental data limit their extrapolative power. This is a major issue in predicting kinematic viscosity, which is a primary measure of alkanes' flow properties, and is defined as a ratio of dynamic viscosity [13] and density. To alleviate this difficulty, we obtain synthetic experimental data through physics based simulations. This would allow neural networks to use the experimental and synthetic data, as well as their discrepancy, to predict the kinematic viscosity of alkanes, and their knowledge is used to model properties of quantum many body systems.

Several different simulation methods are commonly applied to model the properties of physical many body systems, the most famous of which are:

- Monte Carlo simulations [98][46]
- Density functional theory [102]
- Molecular dynamics [60]

Monte Carlo simulations are used to estimate the ground state energy of quantum systems (QMC) [98] and to calculate the physical properties of classical systems by sampling from the Boltzmann distribution (MC) [46]. In density functional theory (DFT), one models the properties of quantum many body systems, just like in QMC. However, the primary objects of study are spatially dependent functions of electron density.

Despite their success in modelling the various properties of quantum and classical systems, neither the QMC nor DFT are commonly used to model the kinematic viscosity of alkanes.

While their use of atomic electronic structure to study many body systems adds a layer of physical complexity, it makes simulations prohibitively computationally expensive to obtain enough synthetic data. MC does not suffer from the same problem, yet "since MC only depends on the positions of the atoms, dynamic information is lacking, and MC cannot be used to e.g. estimate transport properties or diffusion constants" [81].

1.4.2 Introduction to molecular dynamics

By treating atoms as single particles with no underlying substructure, and with the ability to incorporate dynamic information, molecular dynamics does not suffer from the same pitfalls as QMC, DFT, or MC. Despite modeling physical systems at a lower level of physical complexity, molecular dynamics has found applications in many areas of science, from biology [53] to materials science [116]. Motivated by classical statistical physics, the fundamental object of study in molecular dynamics is the classical many body Hamiltonian

$$H^0(\{\vec{p}_i\},\{\vec{r}\}) = \sum_{i=1}^{N} \frac{||\vec{p}_i||_2^2}{2m_i} + U(\{\vec{r}_i\}), \tag{1.8}$$

where H^0 is the Hamiltonian, $\{(r_i, p_i, m_i)\}$ a set of atomic coordinates, momenta and masses, and $U(\{\vec{r}_i\})$ the potential energy. The atomic equations of motion are given by the Hamiltonian dynamics

$$\begin{aligned} \vec{v}_i^0 &= \nabla_{\vec{p}_i} H^0(\{\vec{p}_i\},\{\vec{r}_i\}) = \frac{\vec{p}_i}{m_i} \\ \vec{a}_i^0 &= \frac{-\nabla_{\vec{r}_i} H^0(\{\vec{p}_i\},\{\vec{r}_i\})}{m_i} = \frac{\vec{F}_i}{m_i}, \end{aligned} \tag{1.9}$$

where $(\vec{v}_i, \vec{a}_i)$ symbolize atomic velocity and acceleration, and the force on each atom is given by the gradient of potential energy, known as a force field [35], with respect to atom's position

$$\vec{F}_i = -\nabla_{\vec{r}_i} U(\{\vec{r}_i\}). \tag{1.10}$$

Force field is expressed as a sum of interatomic angle dependent bond energy, the interatomic distance dependent Van der Waals energy, and the interatomic distance dependent Coulomb energy

$$U(\{\vec{r}_i\}) = U_\Theta(\{\vec{\theta}_i\}) + U_{VdW}(\{r_i\}) + U_C(\{r_i\}), \tag{1.11}$$

where $\{\vec{\theta}_i\}$ is the set of angles and $\{r_i\}$ a set of distances between the atoms. The full details of interactions between the atoms are not known, so they are estimated, usually semi-empirically by fitting a specific functional form to experimental data. Some of the most common force fields obtained in such a way are Amber [1], GROMOS [3], CHARMM [2], and OPLS [58]. There has, however, been an increase in the use of machine learning [122] and density functional theory methods [55], [84] in designing force field with the intention of bridging the gap in the accuracy between molecular dynamics and the quantum mechanics based simulation methods such as DFT or QMC.

In this Thesis, we use the SciPCFF force field, which is a Scienomics[2] implementation of the PCFF [118] force field with primarily COMPASS [117], and several proprietary parameters. This force field uses a 9-6 Lennard-Jones (LJ) potential for the Van der Waals potential, and we apply a dielectric constant of 1 to the Coulomb potential. We calculate the Lennard-Jones and Coulomb interactions through the particle mesh algorithm [51] with a precision of 0.0001, and apply a 12Å cutoff without smoothing to both of them. Additionally, we apply tail corrections to the r^{-6} part of the Lennard-Jones, and Coulomb potentials. These corrections do not influence the properties studied in this Thesis, yet affect the pressure, whose average simulation value is used as a check that density simulations have been performed correctly.

1.4.3 Density simulation details

To model density with molecular dynamics, we first prepare simulation input files in the Scienomics' MAPS platform by building a cubical unit cell with a side length of 40Å and density of $800\frac{g}{l}$ at the simulation temperature, and then apply the periodic boundary conditions. Cell's geometry is optimized through 500 steps of conjugate gradient [49] energy minimization to find the best initial simulation conditions.

We perform the simulations in the LAMMPS software [107], and the NPT ensemble for 1ns with a time step of 1fs, to approximate real life experimental conditions. Equations of motion are the non-Hamiltonian generalisations [121] of Equation 1.9 and include the equations for the heat bath, with their full details presented in Shinoda et al [59]. To ensure the system is kept at constant pressure and temperature, we use a 10fs temperature damping and a 350fs pressure damping. During the simulation, a volume measurement is taken every 1ps. The mean and the uncertainty in simulation density are calculated from the measurements after 30ps, after the system has equilibriated (Figure 1.4).

[2]https://www.scienomics.com/

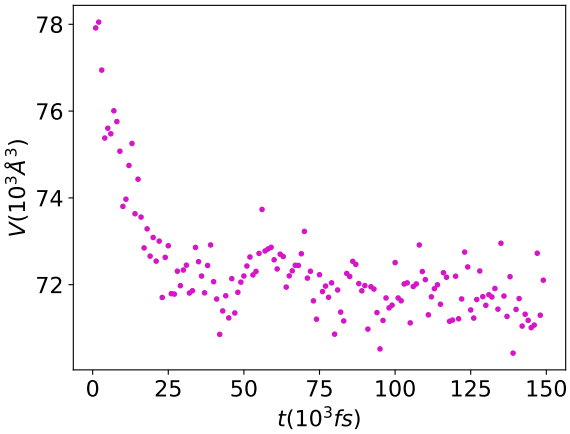


Fig. 1.4 Volume vs simulation time for density of 4-ethyl-4-methylheptane at 25°C for first 150ps of the simulation. $V(t=0) = 64nm^3$ and is not shown since it's a guess system volume. System equilibriates after approximately 25ps, but we conservatively consider volume measurements only after 30ps.

1.4.4 Non-equilibrium molecular dynamics

Molecular dynamics can also be utilized to model dynamic viscosity. A common approach is to use the equilibrium molecular dynamics in the canonical (NVT) ensemble. In this approach, viscosity is calculated through the mathematically exact Green-Kubo relationship [37]

$$\mu = \frac{V}{k_B T} \int_0^\infty \mathbb{E}[P_{xy}(t) P_{xy}(0)] dt, \tag{1.12}$$

where V is the system volume, T the absolute temperature, k_B the Boltzmann constant, $\mathbb{E}$ the expectation operator, and P_{xy} the xy component of the viscous stress tensor, given by

$$P_{xy} = \sum_{i=1}^{N} \frac{p_i^x p_i^y}{m_i} - \sum_{i>j} (x_i - x_j) \frac{\partial U_{ij}}{\partial y_j}, \tag{1.13}$$

where i is the atomic index, p^j the j^{th} component of atomic momenta, x and y atomic coordinates, and U_{ij} the potential energy arising from the interaction between two atoms. Since the viscous stress autocorrelation function ($\mathbb{E}[P_{xy}(t)P_{xy}(0)]$) effectively vanishes for large times, the integral in Equation 1.12 can be truncated without compromising the accuracy of the Green-Kubo method. However, since it vanishes very slowly for larger molecules, increasing by an order of magnitude when several carbon atoms are added to a molecule [76], equilibrium molecular dynamics is impractical to use for larger alkanes.

One can, however, perform the simulations outside of equilibrium to model viscosity of heavy alkanes. We accomplish this by adding a term $H^{'}$ to the Hamiltonian in Equation 1.8. In the case of dynamic viscosity, this term takes the form of Doll's tensor [124]

$$H^{'} = \sum_{i=1}^{N} \vec{r}_i \otimes \vec{p}_i : (\nabla \vec{u})^T, \tag{1.14}$$

where $\otimes$ symbolizes the outer product, : the dyadic product, T a matrix transpose, and $(\nabla \vec{u})^T$ the strain tensor whose elements are given by the partial derivative of speeds with respect to coordinates. The modified, DOLLS equations of motion are then

$$\begin{aligned} \vec{v}_i &= \vec{v}_i^0 + (\nabla \vec{u})\vec{r}_i \\ \vec{p}_i &= \vec{p}_i^0 - (\nabla \vec{u})^T \vec{p}_i, \end{aligned} \tag{1.15}$$

where $\vec{v}_i^0$ and $\vec{p}_i^0$ are given by Equation 1.9. For small perturbations from the equilibrium, these equations of motion agree with the linear response theory, and in the limit of a zero strain rate perturbation the Green-Kubo (Equation 1.12) expression for the dynamic viscosity. Still, viscosity simulations are unreliable and inaccurate at small perturbations, and the DOLLS equations of motion cannot describe the full range of shear flow in a non-linear regime [37].

We can, however, adapt the DOLLS equation of motion for the momenta, with the only difference the disappearance of the transpose of the strain rate

$$\vec{p}_i = \vec{p}_i^0 - (\nabla \vec{u})\vec{p}_i. \tag{1.16}$$

These equations of motions are known as the SLLOD equations of motion [38], and they lead to the same results as the DOLLS formalism in the linear regime, but can accurately describe the shear flow for all the shear rates [37].

1.4.5 Viscosity simulation details

With the SLLOD equations of motion, we can perform the non-equilibrium molecular dynamics viscosity simulations. We do so by simulating the Couette flow, with the shear applied only in the xy plane, such that $(\nabla \vec{u})_{xy} = \dot{\gamma}$ is the only non-zero strain rate element. To physically reproduce the Couette flow in the simulation box, we impose the Lees-Edwards boundary conditions, which, unlike the periodic boundary conditions for density, alter the x positions and velocities when the particle leaves the simulation box in the y direction [70].

Preparing the viscosity simulations follows the same recipe as in subsection 1.4.3, but a simulation cell has the experimental/best approximation of density to ensure that simulation is performed at an accurate pressure. Simulations are performed in the NVT ensemble, and equations of motion are integrated with the velocity Verlet algorithm

$$\begin{aligned} \vec{r}_i(t+\Delta t) &= \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{1}{2}\vec{a}_i(t)(\Delta t)^2 \\ \vec{v}_i(t+\Delta t) &= \vec{v}_i(t) + \frac{1}{2}\Big(\vec{a}_i(t) + \vec{a}_i(t+\Delta t)\Big)\Delta t, \end{aligned} \tag{1.17}$$

where t is the time, and $\Delta t = 1\,fs$ is the timestep. The simulation cell is kept at a constant temperature through the Nosé-Hoover chain thermostat [83] with a 100fs time damping, while a 0.2 drag coefficient is introduced to prevent the temperature oscillations. Simulation time varies as a function of shear rate to improve the confidence in viscosity predictions (Table 1.1). The *xy* component of the stress tensor (P_{xy}) is recorded every 100fs and its expectation value and uncertainty calculated the same way as for density.

$\log(\dot{\gamma})$ (s^{-1})	T_{sim} (ns)
10.30-12.00	1
10.10-10.30	2
9.35-10.10	4
<9.35	8

Table 1.1 Viscosity simulation times

The expected value of kinematic viscosity at a shear rate $\dot{\gamma}_i$ is calculated as the ratio of the negative expected value of the xy component of the viscous stress tensor and a product of the simulation shear rate and liquid density

$$\eta(\dot{\gamma}_i) = \frac{-\mathbb{E}[P_{xy}]}{\rho\,\dot{\gamma}_i}, \tag{1.18}$$

where $\mathbb{E}$, ρ and $\dot{\gamma}_i$ denote the expectation operator, alkane's density and the shear rate, while its uncertainty is calculated as the ratio of the uncertainty in the xy component of the viscous stress tensor and a product of shear rate and liquid density

$$\delta\eta(\dot{\gamma}_i) = \frac{\delta P_{xy}}{\rho\,\dot{\gamma}_i}, \tag{1.19}$$

where δP_{xy} denotes the uncertainty in the xy component of the shear stress tensor and we have neglected the uncertainty in density, since typically $\frac{\delta\rho}{\rho} \ll \frac{\delta P_{xy}}{P_{xy}}$. Uncertainty in kinematic viscosity can in principle be reduced by increasing the simulation time, but its ultimate minimum is in practice limited by its dependence on the reciprocal shear rate.

Since we cannot perform the simulations at small shear rates, viscosity's shear rate profile is first fitted to the Carreau model

$$\eta(\dot{\gamma}) = \eta_\infty + (\eta_0 - \eta_\infty)[1 + (\lambda\dot{\gamma})^2]^{\frac{n-1}{2}}, \tag{1.20}$$

where η_0 is Newtonian viscosity, η_∞ is the value of lower Newtonian plateau, n a nonnegative parameter that determines the shape of the Carreau curve between two plateaus, and λ determines the range of shear rates between the two plateaus. To calculate Newtonian viscosity, we minimize the weighted least squares (WLS) cost function

$$C(\eta_0, \eta_\infty, n, \lambda) = \sum_i \frac{[\eta_i - \eta(\gamma_i)]^2}{\delta\eta_i^2}, \tag{1.21}$$

where η_i is the simulation result of kinematic viscosity at γ_i,$\eta(\gamma_i)$ is the Carreau model viscosity at shear rate γ_i, and $\delta\eta_i$ is the uncertainty in kinematic viscosity. The choice of cost function ensures that we assign higher weights to viscosity results at higher shear rates, with a larger signal to noise ratio which ensures a more reliable extrapolation of Newtonian viscosity. We minimize the WLS cost function with the Levenberg-Marquardt algorithm [71] [80] and the initial parameter guesses of

$$\{\eta_0, \eta_\infty, n, \lambda\} = \left\{\max\{\eta\}, \min\{\eta\}, 1, \frac{1}{\min\{\dot{\gamma}\}}\right\}, \tag{1.22}$$

where $\{n\}$ and $\{\gamma\}$ are the set of viscosity simulation results and shear rates at which we perform the simulations.

1.4.6 Data blocking

Since the motion of atoms and molecules during the molecular dynamics simulations is deterministic, consecutive measurements of physical quantities are correlated, which results in underestimating their uncertainty.

To accurately determine the uncertainty in the property of interest, we use data blocking [40]. In data blocking, consecutive measurements are first assembled into blocks of equal size. Next, the mean of each block is taken as its representative value. The uncertainty in the property of interest is calculated as a standard deviation in the mean in the blocked set. To

obtain an actual value of uncertainty, blocking procedure is performed iteratively until the uncertainty reaches its maximum, which indicates that the measurements are decorrelated. In this Thesis, the number of data entries is halved with each blocking round.

We illustrate the data blocking procedure by determining the uncertainty in density of 4,4-dimethyl-heptane at 100°C (Figure 1.5), which arises from expansion and contraction of the simulation cell. In this example, volume measurements become uncorrelated after five blocking rounds (highlighted by a blue dot with error bars in Figure 1.5).

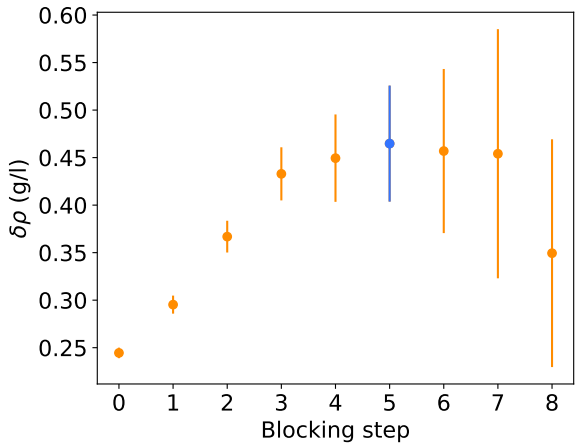


Fig. 1.5 Applying data blocking to determine the uncertainty in density of 4,4-dimethyl-heptane at 100°C. Dots represent expected values of uncertainties while bars represent their uncertainties.

1.5 Mixtures

Base oils are mixtures, comprising several types of alkanes and the knowledge of pure alkanes' physical properties is insufficient to computationally design them. This is because different alkane species interact, so their properties cannot be predicted from weighted averages of pure alkanes properties [87]. To resolve this issue, several semi-empirical mathematical expressions have been developed to model properties of alkane mixtures as a function of their components' molar fractions [86] [44] [100] [62] [9]. None of these expressions, however, have enough extrapolative power nor offer new physical insight, so we do not use them in this Thesis.

The lack of experimental physical property data for mixtures of alkanes is the reason why we also do not use the neural networks or molecular dynamics to model properties of mixtures. The smaller amount of experimental data coupled with the larger number of compounds than for pure alkanes mean neural networks cannot at present establish the connection between

the mixtures' composition and their properties, and that sufficient number of molecular dynamics simulations cannot be performed in an allocated time period for this project to expand the mixtures' dataset to a level where neural networks can use it. Instead, we derive a mathematical expression for properties of mixtures whose components are uniformly spatially distributed molecules in a liquid phase (Figure 1.6), and apply it to alkanes. To apply this mathematical expression to mixtures of alkanes, we first fit it to experimental data and then use the best fit parameters to predict properties of new alkane mixtures.

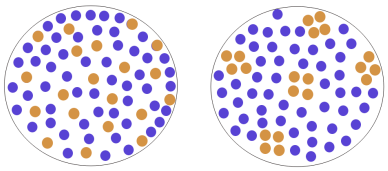


Fig. 1.6 Examples of two mixtures with different spatial molecular distributions. The mixture on the left comprises molecules that follow a uniform spatial distribution, while the ones on the right form clusters with the other molecules of the same type.

1.6 Outline

With the motivation for studying the properties of alkanes and the basics of methods used to study them in place, we now outline the body of this Thesis:

- In chapter 2, we encode molecular structure of linear, single-branched, and double-branched alkanes into a set of non-negative integers, and use neural networks to correlate it to flash point, density and viscosity of linear alkanes, and the boiling point, molar heat capacity, melting point, and vapor pressure of alkanes with up to 12 carbon atoms and 2 branches.
- In chapter 3, we model the density of alkanes with up to ten carbon atoms with molecular dynamics, present a statistical scheme that corrects the systematic error in simulation results, develop a sampling algorithm that automatically selects the shear rates at which to perform the viscosity simulations, and apply this sampling algorithm to linear alkanes.
- In chapter 4, we use the neural networks to model the molecular dynamics density values and extrapolate to density of 11-heptyltricosane, 8,11-dipentyloctadecane, and 8,14-dipentylhenicosane at 40°C and 100°C to obtain viscosity simulation state points,

then perform the viscosity simulations with the help of the sampling algorithm from the previous chapter, and use the simulation results to calculate the viscosity index.

- In chapter 5, we develop a theory of mixtures of molecular liquids whose molecules follow a uniform probability distribution, and apply it to model the molar volume, surface tension, isentropic compressibility, and dynamic viscosity, as well as predict the dynamic viscosity of mixtures of alkanes

In each chapter, we compare our results to experiments, and offer the final remarks in chapter 6.

Chapter 2

Predicting physical properties of pure alkanes with neural networks

We encode alkanes' molecular structure into a set of five non-negative integers to model the boiling point, vapor pressure, and molar heat capacity of light alkanes with up to 2 branches, and flash point and kinematic viscosity of linear alkanes with artificial neural networks. The accuracy of neural network models is assessed through cross-validation and compared to linear regression and several other semi-empirical methods. Neural networks show higher accuracy, consistency and confidence than all the models, suggesting they can predict the boiling point, the vapor pressure and heat capacity of alkanes. However, due to insufficient amount of branched alkane data, they cannot predict the flash point or kinematic viscosity, while accounting for additional physical effects is necessary to predict their melting point.

2.1 Introduction

2.1.1 Motivation

Lubricants are an important component in modern industry. They are used to reduce friction between surfaces, protect them from wear, transfer heat, remove dirt, and prevent surface corrosion to ensure the smooth functioning of mechanical devices. The demand for lubricants makes them an important economic component in the oil and gas business, while their importance is only expected to grow, even as we move towards a future where fossil fuels are a less significant source of energy.

A typical industrial lubricant comprises mainly base oils, which are a mixture of predominantly alkanes with between 18 and 40 carbon atoms, and base oil enhancing additives. Unfortunately, the connection between the lubricants' composition and their properties is not understood. Understanding the link between the physical properties of pure alkanes and their molecular structure is the first step in illuminating this connection. This would be a first step towards computationally designing optimal base oils, which would motivate the distillation of base oil constituents to approach this optimum in practice.

Melting, boiling, and flash point, molar heat capacity, vapor pressure, density, and dynamic viscosity are key alkanes' physical properties for application in base oils. Boiling and melting point tell us the range of temperature values at which alkane is a liquid. Dynamic viscosity and density are usually not useful individually, yet their ratio, kinematic viscosity is a primary measure of alkane's internal resistance to flow. Vapor pressure is a measure of alkane's tendency to evaporate, and flash point of its flammability in the presence of fire, so are key indicators of alkane's safety.

2.1.2 Previous work

The physical properties of alkanes relevant for base oil lubricant design have previously been modeled with a variety of semi-empirical and empirical methods. Wei explored the relationship between rotational entropy and the melting point [60], while Burch and Whitehead used a combination of molecular structure and topological indices to model the melting point of single branched alkanes with fewer than 20 carbon atoms [15]. To predict the normal boiling point of alkanes, Messerly et al. merged an infinite chain approximation and an empirical equation [88], while Burch, Wakefield, and Whitehead [16] used topological indices and molecular structure to model it for alkanes with fewer than 13 carbon atoms and Constantinou and Gani [26] developed a novel group contribution method to calculate it for various organic compounds.

Mathieu developed a group contribution based method to calculate the flash point of various alkanes [36], while Ruzicka and Domalski estimated the heat capacity of various liquid alkanes using a second order group additivity method [112]. De La Porte and Kossack have developed a model based on free volume theory to study long chain linear alkane viscosity as a function of temperature and pressure [68], Riesco and Vesovic have expanded a hard sphere model to study similar systems [99], and Novak has established a corresponding-states model to study viscosity of linear alkanes for the entire fluid region [75]. Marano et al. developed an empirical set of asymptotic behavior correlations to predict the physical properties of a limited family of alkanes and alkenes [77],[78],[79]. Alqaheem and Riazi, and Needham et al. have explored correlations between different properties ([6],[96]) to predict the missing values.

While all the semi-empirical and empirical approaches have merits they do not possess enough extrapolative power and cannot be used to predict the physical properties of alkanes. Fortunately, there is another statistical tool that we could use to accomplish this goal, called an artificial neural network (ANN) [14],[57] (Figure 1.2). Their ability to accurately approximate continuous functional relationships between the variables [31] [106] made them a cornerstone of the digital age, with applications from computer vision to digital marketing, and they have already been applied to physical properties of organic compounds in a variety of settings. Suzuki, Ebert and Schüürmann used physical properties and indicator variables for functional groups to model viscosity as a function of temperature for 440 organic liquids [119], and Ali implemented a conceptually similar approach to model vapor pressure as a function of temperature for various organic compounds [120]. Hosseini, Pierantozzi, and Moghadasi, on the other hand used pressure, pseudo-critical density, temperature and molecular weight as neural network inputs to model dynamic viscosity of several fatty acids and biodiesel fuels as a function of temperature [54]. In this chapter we apply the neural network protocol from section 1.3, since the neural network outputs are property uncertainties as well as their average values.

2.1.3 Molecular basis

To capture the connection between alkanes' physical properties and their molecular structure, we develop molecular basis, which uniquely encodes the structure of every linear, single-branched and double-branched alkane into a set of five non-negative integers up to optical isomerism, which has no effect on the properties studied here. This allows us to have a set of input nodes for each property without possessing experimental alkane data. Molecular basis of an alkane is first found by representing a molecule as a graph (Figure 2.1) and then determining its elements as

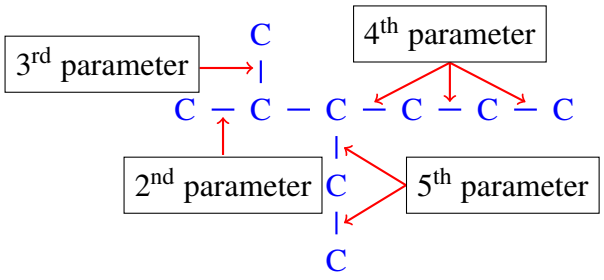


Fig. 2.1 The molecular basis of 3-ethyl-2-methylhexane comprises the five parameters (9,1,1,3,2).

1. The number of carbon atoms.
2. The smaller number of C-C bonds between the end of the longest carbon chain and its closer branch.
3. The number of C-C bonds in the branch closer to an end of the longest carbon chain.
4. The number of C-C bonds between the other end of the longest carbon chain and its closer branch.
5. The number of C-C bonds in the second branch.

Note that if an alkane has a single branch, the last two basis elements are 0, while if an alkane is linear, only the first element is nonzero.

2.2 Results

2.2.1 Simulation protocol

In this section, we apply artificial neural networks to model the physical properties of alkanes. In subsection 2.2.2, we model the molar heat capacity at 25°C, in subsection 2.2.3 the boiling point, in subsection 2.2.4 the vapor pressure, in subsection 2.2.5 the flash point, in subsection 2.2.6 the melting point, and finally in subsection 2.2.7 the kinematic viscosity. The predictive power of neural networks is estimated through cross-validation (subsection 1.3.3), and the accuracy of neural network models compared to linear regression (LR) models with the same independent variables and the same cross-validation data split, as well as to another semi-empirical method where possible. We assess the models' accuracy through the coefficient of determination R^2 and the average absolute deviation (Δ)

$$R^2 = 1 - \frac{\mathbb{E}[(Y_{\text{mod}} - Y_{\text{exp}})^2]}{\sigma^2[Y_{\text{exp}}]}$$
$$\Delta = \mathbb{E}[|Y_{\text{mod}} - Y_{\text{exp}}|], \tag{2.1}$$

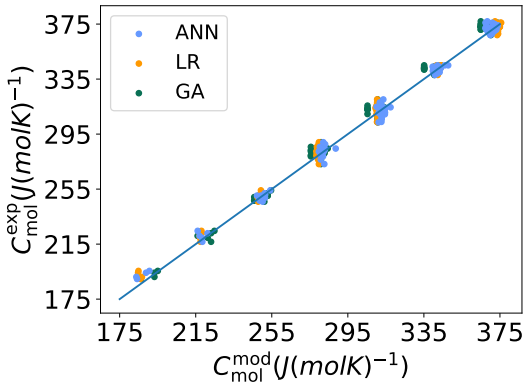
where Y_{mod} symbolizes model estimate of the experimental physical property values Y_{exp}, $\mathbb{E}$ the expectation, and σ^2 the variance operator. Coefficient of determination has been chosen for its frequent use in linear regression, while Δ allows us to directly assess the closeness of model to experimental values. To assess the models' performance further, we also use the standard deviation in the absolute deviation (σ_Δ) and the expected model uncertainty (δ)

$$\sigma_\Delta = (\sigma^2[|Y_{\text{mod}} - Y_{\text{exp}}|])^{0.5}$$
$$\delta = \mathbb{E}[\delta Y_{\text{mod}}], \tag{2.2}$$

where δY_{mod} is the uncertainty in the modelled property values. These two additional metrics help us assess the consistency in model accuracy and the confidence in their predictions.

2.2.2 Heat capacity

We model the molar heat capacity of branched alkanes with 5 to 12 carbon atoms in a liquid phase at 25°C, whose 175 experimental values taken from the TRC Thermodynamic Tables [4]. To do so, we train neural networks with 6 hidden nodes and molecular basis as inputs, and perform 5-fold cross-validation. We also compare model accuracy with a second order group additivity (GA) method [112]. While all three methods accurately model heat capacity, the neural network is the most accurate, with the highest accuracy, results' consistency, and the smallest uncertainty, and the group contribution method is the least accurate, with the lowest accuracy and no uncertainty in predictions (Figure 2.2).



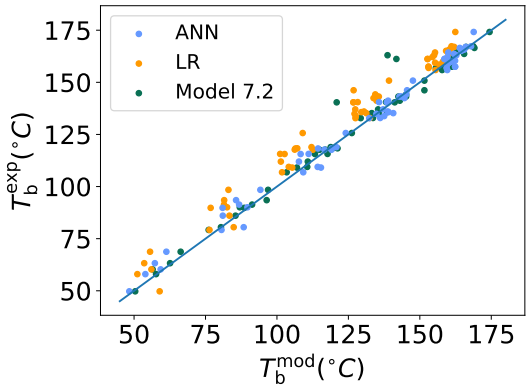
Method	R^2	Δ	σ_Δ	δ
ANN	0.996	2.20	2.25	0.92
LR	0.995	2.37	2.25	7.64
GA [112]	0.994	2.87	2.89	N/A

Fig. 2.2 Parity plot of experimental vs neural network, linear regression, and group additivity [112] heat capacity values are presented on the left, and the statistical summary on the right. Δ, σ_Δ, and δ all have the units of $\frac{J}{Kmol}$. Uncertainties have been omitted to enhance the clarity of presentation.

2.2.3 Boiling point

Next, we model the boiling point of 186 linear, single-branched and double-branched alkanes with 5 to 12 carbon atoms, whose experimental values have been obtained from the TRC Thermodynamic Tables ([4]). We train neural networks with 6 hidden nodes and perform a 5-fold cross-validation. For 63 of these alkanes, we compare results to model 7.2 from Burch et al [16].

The neural network models have the highest R^2 and the lowest σ_Δ, but model 7.2 has the lowest Δ (Figure 2.3). Model 7.2 performed badly on 3-ethyl-2-methylheptane, 3-ethyl-3-methylpentane and 3-ethyl-3-methyl pentane, with the deviations of 19.4°C, 19.5°C and 24.3°C, and R^2 penalizes the outliers more heavily than Δ. Moreover, despite a low Δ, results obtained from model 7.2 are direct fits to data and not their cross-validation extrapolations, so no metric is indicative of model 7.2's predictive power. Consequently, we deduce that neural networks offer superior performance. Across a full dataset, neural networks outperform linear regression on all metrics, with a higher R^2 (0.993 vs 0.958), and a smaller Δ (2.95 °C vs 7.15 °C), σ_Δ (3.55 °C vs 8.82 °C) and δ (1.41 °C vs 16.01 °C).



Method	R^2	Δ	σ_Δ
ANN	0.987	2.90	2.23
LR	0.923	7.77	4.14
Model 7.2 [16]	0.977	2.47	4.37

Fig. 2.3 Parity plot of experimental vs neural network, linear regression, and model 7.2 [16] boiling point values are presented on the left, and the statistical summary on the right. Δ and σ_Δ have the units of °C. Uncertainties have been omitted to enhance the clarity of presentation.

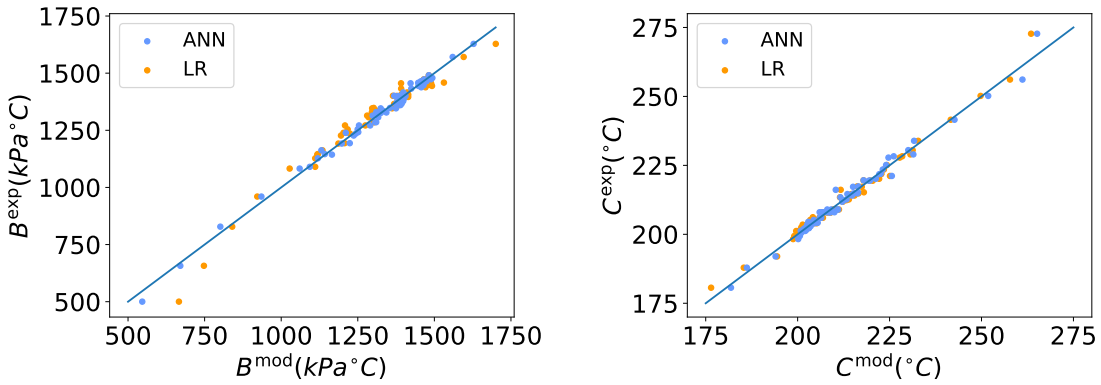
2.2.4 Vapor pressure

Vapor pressure is rarely experimentally recorded directly; instead, it is first measured as a function of temperature, fit to a semi-empirical Antoine equation

$$\log_{10} p = A - \frac{B}{C+T}, \tag{2.3}$$

and the coefficients A, B, C are recorded instead. Coefficients A and B arise from the solution to the Clausius-Clapeyron relation in an ideal gas approximation, while C is empirical and introduced to capture the temperature dependence of latent heat, and temperature T is measured in °C. The Antoine equation is assumed to give an accurate description of alkanes' vapor pressure between the temperatures at which $p = 0.0013kPa$ and $p = 1.97kPa$.

To model the vapor pressure, we first model B and C with neural networks that have 6 hidden nodes on a dataset of 72 alkanes whose values have been obtained from the TRC Thermodynamics Tables [4], and perform a 5-fold cross-validation. While the neural network models outperform linear regression on all the metrics in modeling B, they are of approximately equal accuracy for C, as linear regression has a slightly smaller Δ but a larger uncertainty than the neural network (Figure 2.4). The accuracy of linear regression models for C is an indication that C is an approximately linear function of the molecular basis, and that neural networks have slightly overfit this relationship.



Coefficient	**Method**	R^2	Δ	σ_Δ	δ
B	ANN	0.992	13.04	10.00	112.17
B	LR	0.968	25.85	24.85	6.50
C	ANN	0.984	1.21	1.33	0.495
C	LR	0.984	1.18	1.33	3.72

Fig. 2.4 Parity plot of experimental vs neural network and linear regression. Antoine B coefficient values are presented on the left, while the Antoine C coefficient values are presented on the right. Δ, σ_Δ, and δ have the units of kPa°C for B, and the units of $^{\circ}$C for C. Uncertainties have been omitted to enhance the clarity of presentation.

We can now calculate the vapor pressure and compare results to experimental values. Before calculating the vapor pressure directly, we use the neural network boiling point, B, and C values to calculate A, which is approximately constant for all the alkanes. We have excluded methane from the calculation, as its boiling point calculated from the Antoine equation with the experimental values of the coefficients does not agree with its experimental value, and 3-ethyl-3-methylhexane and 3-ethyl-2-methylhexane as we cannot check that the experimental boiling point value agree with the Antoine equation fit, as they have not been measured.

As vapor pressure is a continuous function of temperature, we cannot use R^2, Δ or σ_Δ to determine the model accuracy. Instead, we define their continuous counterparts, $\tilde{R}^2$, $\tilde{\Delta}$, and $\tilde{\sigma}_\Delta$, given by

$$\begin{aligned}\tilde{R}^2 &= 1 - \frac{\mathbb{E}[\tilde{\mathbb{E}}^2[p_{\text{exp}}(T) - p_{\text{pred}}(T)]]}{\sigma^2[\tilde{\mathbb{E}}^2[p_{\text{exp}}(T) - \bar{p}_{\text{exp}}(T)]]} \\ \tilde{\Delta} &= \mathbb{E}[\tilde{\mathbb{E}}^1[p_{\text{exp}}(T) - p_{\text{pred}}(T)]] \\ \sigma_{\tilde{\Delta}} &= (\sigma^2[\tilde{\mathbb{E}}^1[p_{\text{exp}}(T) - p_{\text{pred}}(T)]])^{0.5},\end{aligned} \tag{2.4}$$

where

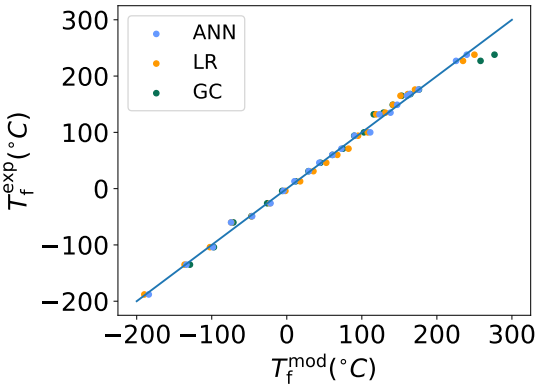
$$\begin{aligned}\tilde{\mathbb{E}}^l[A(T) - B(T)] &= \frac{\int_{T_{\text{min}}}^{T_{\text{max}}} |A(T) - B(T)|^l dT}{T_{\text{max}} - T_{\text{min}}} \\ \bar{p}_{\text{exp}} &= \frac{\int_{T_{\text{min}}}^{T_{\text{max}}} p_{\text{exp}}(T) dT}{T_{\text{max}} - T_{\text{min}}} \\ (T_{\text{min}}, T_{\text{max}}) &= \frac{B_{\text{exp}}}{A_{\text{exp}} - (-1.875, 0.294)} + C_{\text{exp}}.\end{aligned} \tag{2.5}$$

For these three metrics, we obtain $\tilde{R}^2 = 0.986$, $\tilde{\Delta} = 0.028kPa$ and $\sigma_{\tilde{\Delta}} = 0.025kPa$, which shows that neural networks can consistently accurately model the vapor pressure of light alkanes.

2.2.5 Flash point

We model the flash point of linear alkanes with fewer than 31 carbon atoms, whose experimental data was collected from two online sources [109], [105]. The quoted flash point values of henicosane, docosane, tricosane, tetracosane, pentacosane, and hepatcosane are estimated lower bounds rather than their values [105], so we do not use them in.

To model the flash point, we use the number of carbon atoms and boiling point as inputs. The number of carbon atoms is in practice equal to the molecular basis for linear alkanes, and the boiling point has been chosen for its approximately linear correlation with the flash point for hydrocarbons ([6]). We train the neural network models with 5 hidden nodes, perform a 10-fold cross-validation since the dataset is smaller than it was for the other properties, and compare results to the group contribution method [36]. Neural networks perform the best, with the highest accuracy, consistency and the smallest uncertainty, while the group contribution method performs the worst, mispredicting the flash point of octane and triacontane by over 30° C.



Method	R^2	Δ	σ_Δ	δ
ANN	0.998	3.78	3.49	1.66
	0.999*	3.57*	3.57*	1.64*
LR	0.996	6.33	4.20	19.89
	0.995*	6.54*	4.18*	20.69*
GC [36]	0.950	7.60	10.10	N/A

Fig. 2.5 Parity plot of experimental vs neural network, linear regression, and group contribution [36] flash point values are presented on the left, and the statistical summary on the right. Δ, σ_Δ, and δ have the units of °C. Uncertainties have been omitted to enhance the clarity of presentation. $*$ indicates the summary of all results, with methane included, whose flash point the group contribution method does not model.

Since neural network is the most accurate of the three methods, we have used it to predict the flash point values of remaining linear alkanes. These values could replace the erroneous experimental entries and the missing flash point values (Table 2.1).

Name	$T_f^{nn}(^\circ C)$	$\delta T_f^{nn}(^\circ C)$
Henicosane	184.43	7.37
Docosane	194.71	3.60
Tricosane	199.50	5.45
Tetracosane	204.27	8.55
Pentacosane	214.94	6.24
Hexacosane	222.03	8.06
Heptacosane	224.87	6.97
Nonacosane	233.00	3.28

Table 2.1 Neural network predictions of erroneous/missing linear alkane flash point values.

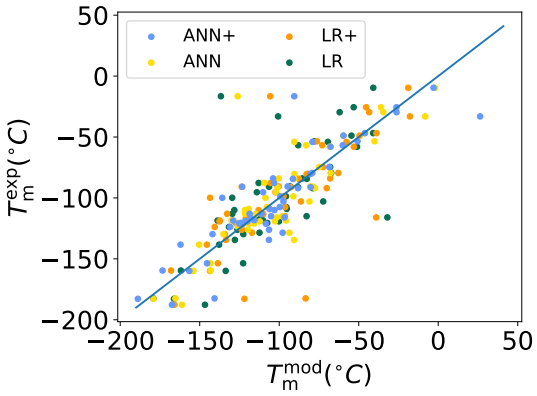
2.2.6 Melting point

We model melting point of 51 alkanes with up to 12 carbon atoms, whose experimental entries have been obtained from TRC Thermodynamic Tables [4]. Initially, we train neural

networks with molecular basis as inputs and 6-hidden nodes, performing an 8-fold cross-validation. Unfortunately, neither the neural networks nor linear regression are accurate, with $R^2 = 0.723$ for the neural network, and $R^2 = 0.528$ for the linear regression.

To improve the accuracy of neural network and linear regression models, we introduce two additional input variables. The first can be either 1 or 0, depending on whether the longest carbon chain has an even or odd number of carbon atoms. Its introduction has been motivated by the empirical observation that alkanes with the even number of carbon atoms in the longest carbon chain have a higher melting point than the ones with an odd number of atoms. The number of molecular symmetries is the second input variable, and its introduction has been motivated by the observation that alkanes with a higher number of molecular symmetries have a higher melting point [60].

With these two additional input variables, we model the melting point again, this time using the neural networks with 8 hidden nodes. The model accuracy has improved for both methods (Figure 2.6), and the neural network models with the additional symmetry information is the most accurate. This shows the usefulness of symmetry information in modeling and signals the importance of entropic effects on the alkanes' melting point.



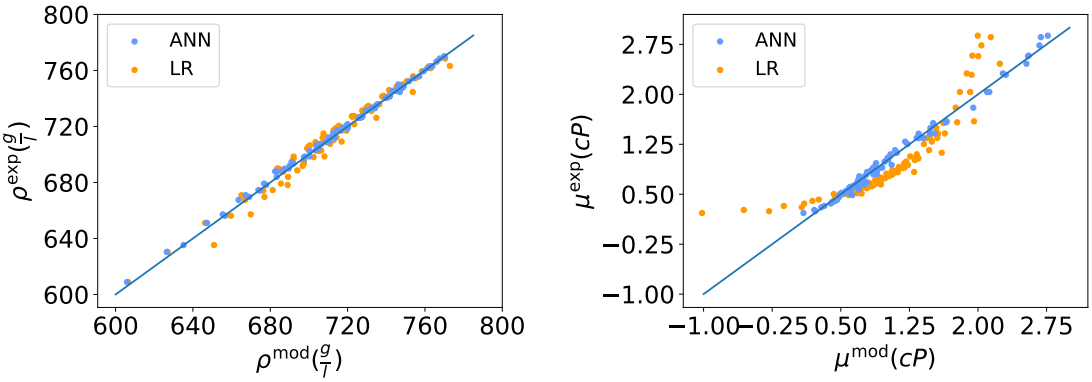
Method	R^2	Δ	σ_Δ	δ
ANN+	0.774	12.71	3.52	14.86
ANN	0.723	14.08	6.27	16.45
LR+	0.573	16.34	50.77	21.34
LR	0.526	18.03	48.91	21.84

Fig. 2.6 Parity plot of experimental vs neural network and linear regression, with and without symmetry inputs melting point values are presented on the left, and the statistical summary on the right. Δ, σ_Δ, and δ have the units of °C. Uncertainties have been omitted to enhance the clarity of presentation.

2.2.7 Kinematic viscosity

Finally, we model liquid density and dynamic viscosity of linear alkanes as a function of temperature. Our dataset comprises 84 data entries for density and 89 for viscosity, with data collected from several research papers ([8], [12], [18], [20], [48], [85], [114], [125]). We

train a set of neural networks for density and another set for viscosity, each with a number of carbon atoms and temperature as inputs, 5-hidden nodes and a 5-fold cross-validation. We note that neural networks perform better than linear regression for both properties on all metrics



Property	**Method**	R^2	Δ	σ_Δ	δ
ρ	ANN	0.997	1.24	1.18	0.92
ρ	LR	0.978	3.91	2.92	5.63
μ	ANN	0.997	1.24	1.18	0.92
μ	LR	0.978	3.91	2.92	5.63

Fig. 2.7 Parity plot of experimental vs neural network and linear regression density values are presented on the left, and the dynamic viscosity values are on the right. Δ, σ_Δ, and δ have the units of $\frac{g}{l}$ for the density, and the units of cP for dynamic viscosity. Uncertainties have been omitted to enhance the clarity of presentation.

Now, we use the neural networks to predict kinematic viscosity of linear alkanes at 20° C, and compare to a model based on a free volume theory (FVT) [68]. Neural networks are shown to be more accurate (Figure 2.8), with a much higher accuracy, results consistency and the uncertainty in predictions which the free volume model does not include, with a higher R^2 (0.982 vs 0.749) and smaller Δ (0.15cSt vs 0.31cSt) and σ_Δ (0.08cSt vs 0.57cSt).

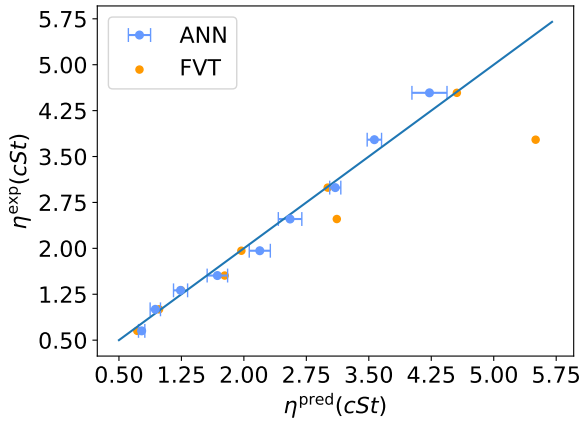


Fig. 2.8 Neural network and free volume theory kinematic viscosity predictions of linear alkanes at 20°C.

2.3 Conclusions

In this chapter, we have used artificial neural networks (section 1.3) to model and predict the physical properties of alkanes from their molecular structure, which was encoded in a molecular basis. We have modelled the boiling point, melting point, heat capacity and vapor pressure of linear, single-branched and double-branched alkanes with up to 12 carbon atoms, and flash point and kinematic viscosity as a function of temperature of linear alkanes. Neural network results have been compared to the ones from linear regression and, where available, to a competing semi-empirical method.

Neural networks perform better than linear regression or other semi-empirical methods, with the highest accuracy, consistency, and confidence in predictions. Still, for molar heat capacity and the Antoine C coefficient, linear regression performs about as well as neural networks. This signals that molecular basis adequately describes the molecular structure for the purposes of modeling these properties, and exhibits a linear relationship with them. A full set of results is presented in Appendix A.

From the applied standpoint, work in this chapter shows that we can use neural networks to predict the boiling point, molar heat capacity, and vapor pressure of heavier alkanes with molecular basis as inputs. The approximately linear relationship between the boiling point and the flash point for linear alkanes offers signs that the same can be done for flash point as well, although the lack of data for branched alkanes prevents us from checking whether the same linear relationship holds for non-linear alkanes as well. Unfortunately, we still cannot predict the melting point, which has not been accurately modeled and for which

more physical effects need to be incorporated, or the kinematic viscosity due to a lack of experimental data for non-linear alkanes. This means that several hurdles still have to be overcome before we can accurately describe the connection between the alkanes' structure and their physical properties.

Chapter 3

Enhancing NEMD with automatic shear rate sampling

We perform molecular dynamics simulations to model density as a function of temperature for 74 alkanes with 5 to 10 carbon atoms and non-equilibrium molecular dynamics simulations in the NVT ensemble to model kinematic viscosity of 10 linear alkanes as a function of molecular weight, pressure, and temperature. To model density, we perform simulations in the NPT ensemble and apply correction factors to exploit the systematic error in the SciPCFF force field. Comparing results to experimental values, we obtain an average absolute deviation of 3.4 $\frac{g}{l}$ *at 25°C and of 7.2* $\frac{g}{l}$ *at 100°C. We develop a sampling algorithm that automatically selects good shear rates at which to perform viscosity simulations in the NVT ensemble and use Carreau model with weighted least squares regression to extrapolate Newtonian viscosity. Viscosity simulations are performed at experimental densities and show an excellent agreement with experimental viscosities, with an average percent deviation of -1% and an average absolute percent deviation of 5%. Future plans to study and apply the sampling algorithm are outlined.*

3.1 Introduction

Alkanes are of great interest to both the academic community and a large number of scientists and engineers using them in industry. In academia, their chemical simplicity makes them an ideal testing ground for applications of novel computational methods in studying physical properties of complex fluids. In industry, understanding alkanes and their properties is essential to produce superior oil and gas products.

Kinematic viscosity, which is a measure of their flow properties, is one of alkanes' most important properties. However, the viscosity of pure alkanes is still poorly understood. While many viscosity measurements of mixtures are made in industrial laboratories on a daily basis, the viscosity of only about 20 pure alkanes has been published in the academic literature, and difficulties in separation of different isomers beyond dodecane prevent engineers and scientists from making measurements of viscosity of large alkanes. Consequently, several theoretical and computational methods have been developed to investigate alkanes' viscosity variation with molecular structure, temperature, and external pressure. For example, De La Porte and Kossack modelled viscosity of long chain n-alkanes with a model motivated by the free volume theory [68]; Riesco and Vesovic used a hard sphere model to predict viscosity of similar systems [99], and Novak modelled viscosity of alkanes with a corresponding states model [75]. Modern statistical methods have also been used to model viscosity of alkanes. Santak and Conduit modelled kinematic viscosity of n-alkanes with a neural network that can make predictions on sparse datasets [113]; Suzuki et al. utilized fully connected neural networks to model viscosity as a function of temperature of various organic compounds [119], while Hosseini et al. used a neural networks and a hard sphere model to model similar systems[54].

Equilibrium molecular dynamics (EMD), frequently applied to model viscosity of light alkanes, is another popular computational method. Cui et al. modelled viscosity of hexadecane, tetracosane, and decane [28], and compared molecular and atomic formalisms for EMD simulations of decane [27]; Singh, Payal et al. modelled viscosity of hexadecane with several force fields [104]; Zhange and Ely modelled viscosity of alkane systems and alcohols [129], while Kondratyuk modelled viscosity of triacontane [67]. Furthermore, Kioupis and Maginn modelled viscosity of a hexane/hexadecane mixture [64], and determined the viscosity number in addition to investigating viscosity variation with pressure of three distinct poly-α-olefins [65] [66], while Mundy et al. predicted viscosity of n-decane, n-hexadecane, 6-pentylundecane, 7,8-dimethyltetradecane, 2,2,4,4,6,8-heptamethylnonane, n-triacontane and squalane [94] and determined pressure-viscosity coefficient of decane [95].

Nevertheless, none of the semi-analytical methods, the modern statistical methods and EMD have been certified to reliably model the viscosity of all alkanes. Semi-analytical

methods do not possess enough sufficient predictive power to be judiciously extrapolated to alkanes outside of the training set, which usually comprises a limited set of light alkanes. Modern statistical methods possess greater extrapolative power than their semi-analytical counterparts, yet their utility is still limited by the lack of experimental data. EMD can in principle be used for all alkanes, but because of slow relaxation of the stress-stress autocorrelation function [101] [28] for larger molecules [76], it is recommended to primarily use it to model viscosity of light molecules [76].

Another physics based simulation method that has gained momentum in the past several decades is the non-equilibrium molecular dynamics (NEMD) [92], in which shear is applied to a molecular system, usually at fixed temperature and volume. A molecular dynamics simulation is performed at several shear rates, and the shear rate profile of the kinematic viscosity is then extrapolated to Newtonian viscosity. In addition to applying EMD, Kioupis and Maginn also used NEMD to model viscosity of hexane/hexadecane binary mixture [64], and of three poly-α-olefins [65] [66], while Mundy et al. utilised NEMD to study viscosity of decane [93] and several large branched alkanes [94]. Cui et al. used NEMD to model viscosity of decane at 25°C, hexadecane at 27°C and 50°C, tetracosane and 10-hexylnonadecane at 60°C, and squalane at 39°C and 99°C [29] [28]; McCabe, Pan, and Evans modelled viscosity of decane [101] [32]; Liu et al. modelled viscosity of squalane and 1-decene-trimer [74]; Cho, Jeong, and Buig modelled viscosity of polymer melts [23]; Yang, Pakkanen, and Rowley determined viscosity index of various lubricant size molecules [128], as well as of several small alkane mixtures [127]; Liu et al. determined a pressure viscosity coefficient of a 1-decene trimer [73]; Allen and Rowley compared different force fields to model viscosity of small alkanes [5], while Khare, de Pablo and Yethiraj modelled viscosity of hexadecane, docosane, octacosane and 5,12-dipropyl-hexadecane [63] and J.D Moore, S.T Cui, H.D Cochran and P.T Cummings modelled viscosity of C100 [91] .

However, despite its past success in modeling viscosity of some alkanes, the contemporary NEMD approach still suffers from three pitfalls. Any viscosity simulation result carries a systematic error from the force field that determines the motion of atoms and molecules. Also, to perform NEMD simulations at accurate external conditions, the density of the alkane of interest needs to be either experimentally known or accurately modelled with molecular dynamics. Despite possessing more experimental data for density than for viscosity, the density of most alkanes is experimentally unknown, and while molecular dynamics simulation results are frequently used to replace experimental values, they need to be in close agreement with true values to be confidently applies as initial conditions in NVT simulations; otherwise, simulations are performed at a wrong external pressure and viscosity simulation results will carry a large systematic error due to viscosity's pressure dependence. Finally, the reliability

of viscosity simulations decreases, while uncertainty in viscosity simulation results increases with decreasing shear rate, making direct identification of Newtonian viscosity difficult, with its accurate extrapolation dependent on performing the simulations at appropriate shear rates. Currently, no computational method is capable of systematically and automatically selecting good simulation shear rates for any alkane at arbitrary external conditions.

In this chapter, we present two computational techniques that enhance the current NEMD method. Firstly, in section 3.2, we use a data driven approach to correct the discrepancies between the experimental and simulation density values. Then, in section 3.3, we develop a sampling algorithm that automatically selects the good shear rates at which to perform the simulations, and apply it to linear alkanes. Density simulations follow the procedure described in subsection 1.4.3, and viscosity simulations the procedure described in subsection 1.4.4. To assess the agreement between the experimental and modeled values, we use an average absolute deviation (Equation 2.1) for density, and the average ($\Delta_{\%}$) and the absolute average percent deviation ($\Delta_{\%}$) for kinematic viscosity, all the metric chosen for their interpretability and widespread use in the literature

$$\begin{aligned} \Delta_{\%} &= 100 \cdot \mathbb{E}\left[\frac{Y_{\text{mod}} - Y_{\text{exp}}}{Y_{\text{exp}}}\right] \\ \Delta_{|\%|} &= 100 \cdot \mathbb{E}\left[\frac{|Y_{\text{mod}} - Y_{\text{exp}}|}{|Y_{\text{exp}}|}\right]. \end{aligned} \tag{3.1}$$

3.2 Density

We perform molecular dynamics simulations in the NPT ensemble for 74 alkanes at 25°C and 34 alkanes at 100°C, and compare to experimental data from the TRC Thermodynamic Tables [4]. To check the reliability of NPT simulation results, we investigate the average pressures obtained in the NPT simulations. During a typical simulation, pressure varies between -1500atm and 1500atm. However, after averaging within a simulation, pressure varies between -30atm and 30atm, with statistical uncertainty obtained through data blocking (subsection 1.4.6) from between 20atm and 30atm, and the atmospheric pressure within the 95% confidence interval, which confirms that density simulations have been performed properly.

Next, we analyse the statistical uncertainty in density simulations. First, we compare uncertainties at two different temperatures. The average uncertainty at 25°C is 0.53 $\frac{g}{l}$, while at 100°C it is 0.61 $\frac{g}{l}$, so there is no indication that increasing the temperature by 75°C increases the statistical uncertainty in density simulation results for small alkanes. Then, we

investigate the uncertainty as a function of molecular weight (Table 3.1) and deduce that increasing molecular weight does not increase the uncertainty in average densities for light linear, single-branched, and double-branched alkanes.

N_C	N_{mol}	$\mathbb{E}[\delta\rho]$ $(\frac{g}{l})$
6	10	0.69
7	16	0.51
8	26	0.57
9	24	0.58
10	32	0.50

Table 3.1 Summary of uncertainty as a function of molecular weight. N_C, N_{mol} and $\mathbb{E}[\delta\rho]$ denote the number of carbon atoms, the number of molecules and the mean value of uncertainty in density.

Initially, we obtain an average absolute deviation of $11\frac{g}{l}$ at 25°C, and of $16\frac{g}{l}$ at 100°C. Since large discrepancies between the experimental and simulated densities lead to inaccurate initial conditions for viscosity simulations if simulation densities are used as inputs, we look to correct them. We do so through a data-driven approach by splitting the alkanes for which we performed simulations into six groups, with each group either a homologous series or a set of homologous series (Table 3.2). The discrepancy between simulations and experiment within each group is approximately constant (Figure 3.1, left) , likely due to the systematic bias in the SciPCFF force field, as the density of alkanes with fewer branches is modelled more accurately, indicating that its parameters for alkanes have been developed mostly from linear alkane data. For each group apart from the linear alkanes, the average group discrepancy is larger at 100°C than at 25°C, possibly due to development of the SciPCFF force field parameters mostly from room temperature data.

Group	N_{mol} (25 °C)	$\lvert\Delta_{25°C}\rvert$ ($\frac{g}{l}$)	N_{mol} (100 °C)	$\lvert\Delta_{100°C}\rvert$ ($\frac{g}{l}$)
linear	5	3.6	5	2.5
methyl series	14	6.1	7	9.4
2,2-dimethyl	5	26	3	34
other dimethyl	28	15	12	22
methyl-ethyl series	14	7.9	4	20
other	8	2.4	3	5.2

Table 3.2 Summary of group discrepancies. At both temperatures, number of molecules and the average absolute deviation and standard deviation in discrepancy are presented. Since the signs of discrepancies are consistent for each group, for all the groups but the linear group, $\Delta = |\Delta|$, with $\Delta = -|\Delta|$ for linear alkanes.

Since viscosity simulations are in general performed at a constant density, results of NPT simulations with large discrepancies are insufficiently accurate to be used as state points for NVT simulations. To obtain more accurate results, we subtract the value of average discrepancy between a group to which an alkane belongs from the simulation result (Table 3.2). A small average pressure variation in simulation results justifies applying the same correction factor at all pressures, since the isothermal compressibility factor is approximately constant for the range of average pressures obtained from simulations [39]. However, since applying correction factors to simulation results is a poor indication of the actual merit of applying them, we perform a leave-one-out cross-validation [57], in which correction factors are calculated from all but one data entry and applied to the remaining data entry, repeating for each entry in a dataset.

After applying a leave-one-out cross-validation, we obtain an average absolute deviation of 3.4 $\frac{g}{l}$ at 25°C (Figure 3.1) and of 7.2 $\frac{g}{l}$ at 100°C, a significant improvement over the results obtained from molecular dynamics simulations. The summary of all the results is presented in Table 3.3, while the parity plot of corrected densities is presented in the right part of Figure 3.1. At 25°C, the model performs the best for linear alkanes and the worst for the 2,2-dimethyl homologous series and the group of other dimethyl alkanes. At 100°C, the model still performs the best for linear alkanes, but now it performs the worst for the methyl-ethyl group, for which the average absolute deviation is 17 $\frac{g}{l}$. Such a large discrepancy arises from a large spread in discrepancies in original simulation results across the ethyl-methyl group. A full list of results can be found in Table B.1 and Table B.2.

Group	N_{mol} (25 °C)	$\lvert\Delta_{25°C}\rvert$ $(\frac{g}{l})$	N_{mol} (100 °C)	$\lvert\Delta_{100°C}\rvert$ $(\frac{g}{l})$
linear	5	0.93	5	0.84
methyl series	14	2.1	7	4.3
2,2-dimethyl	5	5.0	3	5.4
other dimethyl	28	4.9	12	9.6
methyl-ethyl series	14	3.1	4	17
other	8	1.4	3	2.8

Table 3.3 Summary of discrepancies after applying the correction factors and running a leave-one-out cross validation. At both 25 °C and 100 °C, number of molecules and absolute average deviation are presented.

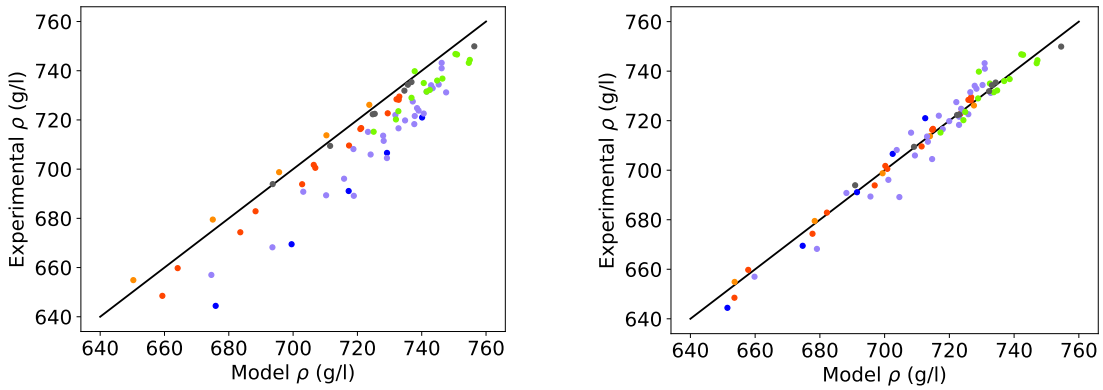


Fig. 3.1 Parity plot of density results vs experimental values before (left) and after (right) correction factors are applied. Orange dots denote the linear alkane series, red dots denote the methyl group, blue dots denote the 2,2-dimethyl series, violet dots denote the group comprising all the other alkanes, light green dots denote the ethyl-methyl group, while the group of all the other molecules is denoted with grey dots.

3.3 Viscosity

3.3.1 Identifying good shear rates

The ratio of speeds due to shear and due to particle interactions is proportional to the shear rate, resulting in a low signal to noise ratio for viscosity simulations performed at low shear rates. For a fixed shear rate, this ratio is smaller for larger temperatures due to smaller relative contribution of the kinetic term to the shear stress tensor, heavier molecules due to inverse relation between speed at a fixed temperature and molecular mass, and at higher pressures due

to an increased virial term contribution arising from closer proximity of molecules at a fixed volume. Therefore, direct identification of Newtonian viscosity with NEMD is challenging, while poor statistics at low shear rates becomes an obstacle in its accurate extrapolation. The range of good shear rates at which to perform viscosity simulations is a priori unknown, and while the authors of previous NEMD studies have performed their simulations at reasonable shear rates, they have selected them manually. Currently, an algorithm to automatically sample good shear rates for an arbitrary alkane at any temperature and pressure does not exist.

To automatically sample good shear rates for an arbitrary alkane, we first run a simulation at the largest shear rate $\dot{\gamma}_0$. Next, we successively decrease the shear rate by a constant $x > 1$ and perform simulations at two smaller shear rates $\dot{\gamma}_1 = \frac{\dot{\gamma}_0}{x}$ and $\dot{\gamma}_2 = \frac{\dot{\gamma}_0}{x^2}$. Then, to assess the vicinity to the upper Newtonian plateau, we calculate the probability that the shear rate profile of kinematic viscosity between two smallest shear rates is concave up, $P[\eta(\dot{\gamma}_2) - \eta(\dot{\gamma}_1) > \eta(\dot{\gamma}_1) - \eta(\dot{\gamma}_0)]$ and compare it to a constant $C \in [0,1]$ under the assumption that kinematic viscosity at each shear rate is normally distributed under its mean and the uncertainty. If $P[\eta(\dot{\gamma}_2) - \eta(\dot{\gamma}_1) > \eta(\dot{\gamma}_1) - \eta(\dot{\gamma}_0)] > C$, we again decrease the shear rate by a constant x and run a simulation at $\dot{\gamma}_3 = \frac{\dot{\gamma}_2}{x}$ before we determine the probability that viscosity's shear rate profile between $\dot{\gamma}_1$ and $\dot{\gamma}_3$ is concave up. The process of performing the simulations at successively smaller shear rates that are a constant fraction of the previous shear rate is repeated until $P[\eta(\dot{\gamma}_n) - \eta(\dot{\gamma}_{n-1}) > \eta(\dot{\gamma}_{n-1}) - \eta(\dot{\gamma}_{n-2})] < C$. To avoid performing simulations with a low signal to noise ratio, we do not perform the simulations at smaller shear. Instead, we perform three more simulations at shear rates uniformly spaced between two smallest shear rates $\dot{\gamma}_{n+1} = \frac{\dot{\gamma}_n + \dot{\gamma}_{n-1}}{2}$, $\dot{\gamma}_{n+2} = \frac{\dot{\gamma}_{n+1} + \dot{\gamma}_n}{2}$, and $\dot{\gamma}_{n+3} = \frac{\dot{\gamma}_{n+1} + \dot{\gamma}_{n-1}}{2}$. In this manuscript, we use $\dot{\gamma}_0 = 10^{12}\mathrm{s}^{-1}$, $x = 3$ and $C = 0.95$ to cover a large range of shear rates with a relatively small number of simulations and continue performing simulations at smaller shear rates only if we're 95% confident that viscosity's shear rate profile in the region of interest is concave up. A flow chart that concisely summarizes the sampling algorithm is shown in Figure 3.2.

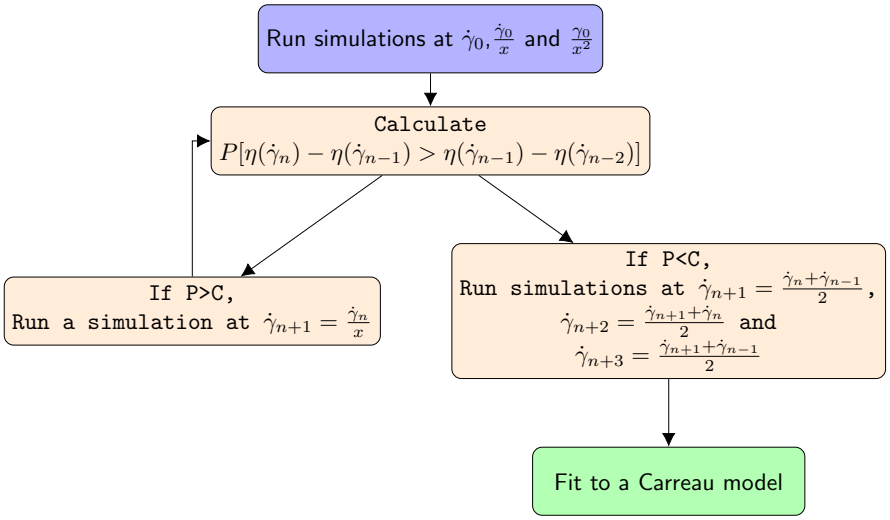


Fig. 3.2 Schematic of the algorithm applied to determine appropriate shear rates.

We illustrate the sampling algorithm in modelling kinematic viscosity of octadecane at 50°C (Figure 3.3). Shear rate is consecutively decreased by a third down to $\log(\dot{\gamma}) = 9.14$, when $\eta = 3.39 \pm 0.29$cSt. Kinematic viscosity at two immediate smaller shear rates ($\log(\dot{\gamma}) = 9.61$ and $\log(\dot{\gamma}) = 10.09$) is simulated and found to be 2.63 ± 0.13cSt and 1.52 ± 0.05cSt. Since $P[\eta(10^{9.14}) - \eta(10^{9.61}) > \eta(10^{9.61}) - \eta(10^{10.09})] = 0.1847$, three more simulations are performed at $\log(\dot{\gamma}) = 9.44$, $\log(\dot{\gamma}) = 9.32$ and $\log(\dot{\gamma}) = 9.54$, and data is fitted to the Carreau model with a WLS regression (section 2.3).

While neither simulation has been performed at a low enough shear rate to directly identify the upper Newtonian plateau, we have extrapolated Newtonian viscosity of 3.24cSt, an excellent agreement with the experimental value of 3.23cSt reported in Caudwell et al. [20].

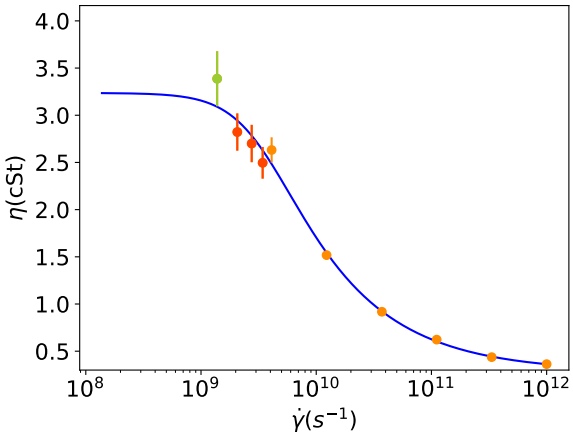


Fig. 3.3 Kinematic viscosity plotted against the shear rate for octadecane at 50°C. Shear rate is plotted on the logarithmic scale. Orange dots represent simulation results at high shear rates; light green dot represents the simulation result at the lowest shear rate, while red dots represent simulation results at intermediate shear rates.

3.3.2 Results

We now study kinematic viscosity of linear alkanes, which serve as a case study for evaluating the reliability and accuracy of the sampling algorithm for two reasons. Firstly, they are the homologous alkane series with the most available experimental data. Secondly, guided by the results of density simulations, we expect the systematic error in the SciPCFF force field for linear alkanes to be small compared to for the other homologous series. Consequently, the discrepancy between simulations and the experiments arises primarily from the remaining imperfections in choosing the shear rates. To directly assess the performance of the sampling algorithm, we perform the simulations at experimental densities.

We first study viscosity as a function of molecular weight, modeling kinematic viscosity of hexane, heptane, octane, nonane, decane, undecane, dodecane, tridecane and tetradecane at 20°C at atmospheric pressure and compare results to experimental values from the TRC Thermodynamic Tables [4] (Table B.3). Simulations accurately reproduce the experimental data, with an average percent error of 5% (Figure 3.4) and the absolute percent error of 6.4%. Simulations are the least accurate for heptane and tetradecane, with the percent errors of 13% and -10%, while experimental values for all the alkanes apart from tetradecane are within the 95% confidence interval. Simulations systematically underestimate kinematic viscosity of decane and heavier alkanes, which we attribute to the small systematic error in the SciPCFF force field that also underestimated the density of linear alkanes.

To further evaluate the performance of the sampling algorithm, we compare the accuracy of our prediction for decane to the prediction made in Cui et al.[29] at 25°C. Our prediction of 1.13±0.08cSt is in excellent agreement with the experimental value of 1.24cSt and compares favourably with their prediction of 0.84±0.11cSt against the experimental value of 1.17cSt.

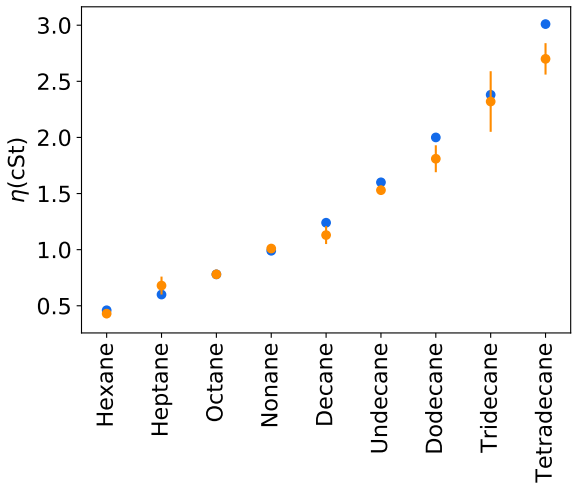


Fig. 3.4 Viscosity of linear alkanes at 20°C. Blue dots present experimental data, while orange dots represent molecular dynamics predictions with accompanying statistical uncertainty.

Then, we explore the variation of viscosity with pressure, with tridecane at 60°C as a case study, and the experimental data coming from Daug et al. [33] (Table B.4). Simulations results are in excellent agreement with experiments (Figure 3.5), with an average percent error of 2%, an absolute percent error of 4%, and the least accurate prediction at 100MPa, with a percent error of 8%. All the experimental values are within a 95% confidence interval of our predictions.

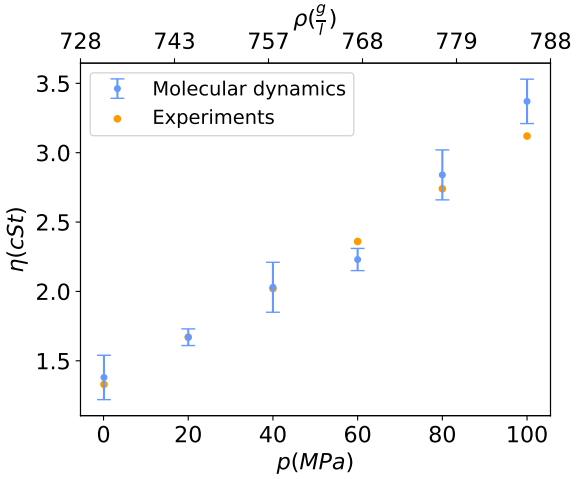


Fig. 3.5 Viscosity of tridecane at 60°C as a function of pressure. Kinematic viscosity varies approximately with the simulation density, highlighting the importance of using the accurate initial simulation conditions.

Next, we calculate the pressure-viscosity coefficient, which is a measure commonly used in industry to assess the pressure gradient of alkanes' viscosity at a fixed temperature T. The pressure-viscosity coefficient appears in the exponent of the following equation:

$$\eta(p,T) = \eta_{\text{atm}}(T)\mathrm{e}^{\alpha \mathrm{p}}, \tag{3.2}$$

where $\eta_{\text{atm}}(T)$ is a value of kinematic viscosity at atmospheric pressure and the temperature of interest, and p is the pressure. An experimental value of the pressure viscosity coefficient is 0.00886 MPa^{-1}, while the simulations predict 0.00869 MPa^{-1}. A percent error of only -2% and the absolute percent error of 4% further confirm that we can accurately capture the variation of alkane's viscosity with pressure.

Thirdly, we study the variation of viscosity with temperature, focusing on viscosity of octane, dodecane, and octadecane. Simulations are performed at temperatures at least 20°C above alkanes' melting points to avoid the crystallization of the cell. We first model viscosity of octane (Table B.5) and dodecane (Table B.6), whose experimental viscosity's temperature profile was obtained from Caudwell et al. [20] [18]. Simulation results are in excellent agreement with experiments, with the average percent error of -0.4% for octane and of 4% for dodecane, and the absolute percent error of 4% and of 8% for dodecane (Figure 3.6). All the experimental values lie within the 95% confidence interval of mean simulation predictions apart from the octane results at 25 °C and 100 °C and the dodecane results at 200°C, primarily due to an excellent fit of viscosity's shear rate profile to the Carreau model.

Finally, we model viscosity of octadecane, whose experimental values were obtained from Caudwell et al.[18]. We observe that the simulations are in excellent agreement with experimental values (Figure 3.6), with an average percent deviation of 0.4%, and the absolute average percent error of 4%. Viscosity at 100°C was simulated with the smallest accuracy, with a 6% percent deviation, while all the results apart from the one at 200°C are within a 95% confidence interval. The longest total simulation time to model viscosity at a fixed temperature is 36ns, which is only 5.14 times longer than the time spent to model viscosity of hexane at 20°C. Such a small increase in total simulation time gives us further confidence that we can apply the sampling algorithm to heavy alkanes without requiring excessive computational resources like in equilibrium molecular dynamics.

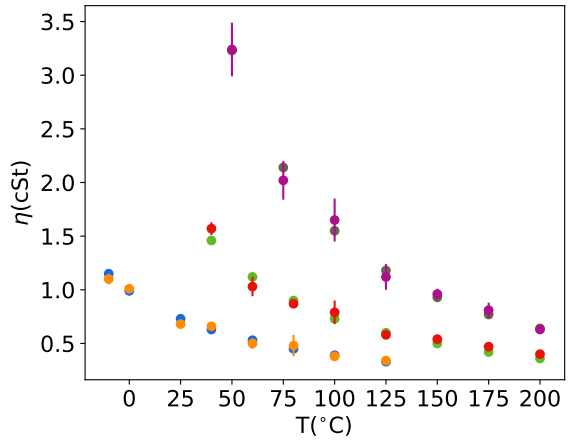


Fig. 3.6 Viscosity of octane, dodecane and octadecane as a function of temperature. Blue, green and grey dots represent experimental values of their viscosity, while orange, red and purple dots with accompanying statistical uncertainty represent values predicted by the NEMD simulations.

With all the molecules' viscosity modeled , we analyse the overall accuracy of viscosity simulations. A parity plot showing experimental values against simulation results for all the alkanes studied is shown in Figure 3.7. Simulations are in excellent agreement with experiments, with an average error of -1% and the average absolute percent error of 5%.

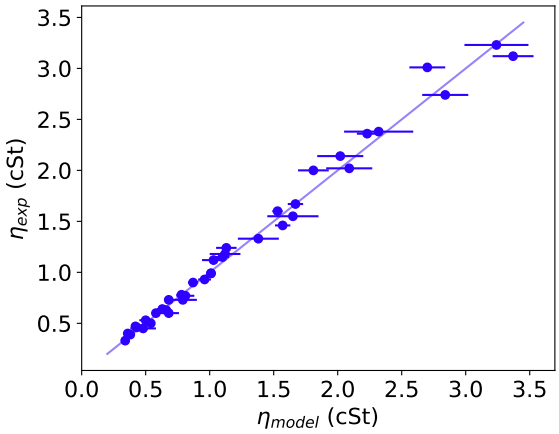


Fig. 3.7 Parity plot showing experimental viscosity values against the NEMD simulation results.

Then, we study the percent error in our models as a function of predicted viscosity to assess whether the simulations perform equally well at all viscosities (Figure 3.8). We note that the average error fluctuates between about -10% and 10% for all the modelled viscosities, indicating that the sampling algorithm could be successfully applied to heavy alkanes.

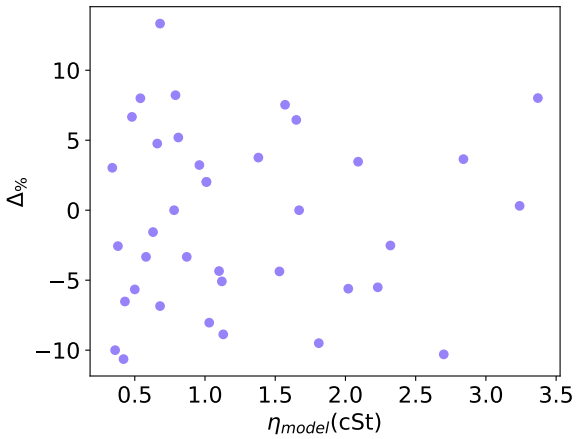


Fig. 3.8 Percent error for all the data as a function of viscosity simulation results.

Finally, we study the statistical uncertainty in the mean predictions. We observe that the uncertainty in NEMD viscosity predictions increases approximately linearly as a function of predicted viscosity (Figure 3.9), with an $R^2 = 0.61$ of the linear fit. The approximate linear dependence of uncertainty on viscosity arises from uncertainty in the best fit parameters' dependence on the matrix of uncertainties in kinematic viscosity at different shear rates, whose entries are inversely proportional to the shear rate.

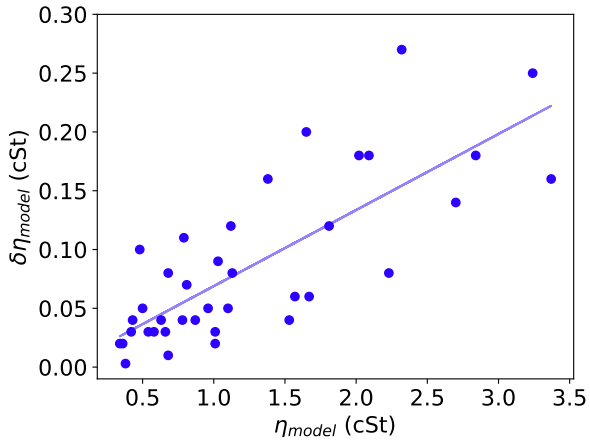


Fig. 3.9 Uncertainty in Newtonian viscosity predictions as a function of viscosity predictions and its best linear fit $\delta\eta_{\text{model}} = 0.065\eta_{\text{model}} + 0.004$

3.4 Conclusion

In this chapter, we modeled density and dynamic viscosity of alkanes with molecular dynamics. For density, we first performed molecular dynamics simulations in the NPT and then applied correction factors to rectify the systematic error arising from the SciPCFF force field, obtaining an absolute deviation of 3.4 $\frac{g}{l}$ at 25°C and an absolute deviation of 7.2 $\frac{g}{l}$ at 100°C. To model viscosity, we have developed a sampling algorithm to identify the good shear rates at which to perform viscosity simulations. We have then utilised this algorithm to study the kinematic viscosity of hexane, heptane, octane, nonane, decane, undecane, dodecane, tridecane, and tetradecane at 20°C; viscosity of tridecane at 60°C as a function of pressure, and viscosity of octane, dodecane, and octadecane as a function of temperature at experimental densities. Simulation results are in excellent agreement with experiments, with an average percent error of -1% and the average absolute percent error of 5%. The average percent error stays approximately constant and fluctuates about 10% in magnitude as a function of viscosity, while the uncertainty in viscosity predictions increases approximately linearly with increased viscosity.

Work presented in this chapter sets a solid foundation for performing simulations of more complex alkanes. The data driven approach to correct densities is ad hoc, and offers no novel conceptual insights, yet upon performing more density simulations, more patterns in the discrepancy between the simulations and the experiments can be identified, so better correction factors to simulation densities can be applied. This would ensure that good initial conditions for viscosity simulations are used when experimental densities for alkanes of

interest are not available. Beyond automatically identifying good shear rates, our sampling algorithm can be straightforwardly applied in high throughput screening, and its generality means that it can be used as a basis to study viscosity of other liquids with a known functional dependence on shear rates. To make the sampling algorithm better, one can also study its mathematical properties, which would help the non-equilibrium molecular dynamics simulations in the NVT ensemble to reach their optimum.

Chapter 4

Predicting kinematic viscosity and its index of three heavy alkanes

We perform non-equilibrium molecular dynamics simulations in the NVT ensemble to predict kinematic viscosity of 11-heptyltricosane, 8,11-dipentyloctadecane, and 8,14-dipentylhenicosane at 40°C and 100°C, and calculate their viscosity index. To determine viscosity simulation state points, we perform molecular dynamics density simulations at 25°C and 100°C for 141 linear, single-branched or double-branched alkanes with between 10 and 41 carbon atoms, apply neural networks to extrapolate to other molecules and linearly interpolate network predictions to 40°C. Neural network models are in excellent agreement with the molecular dynamics density values, with cross-validation $R^2 \geq 0.996$*, but viscosity simulations are not in a good agreement with the experiments, with a percent error up to 55%. We identify the systematic error in the SciPCFF force field as the main source of discrepancy between simulations and the experiments.*

4.1 Introduction

Viscosity is a measure of alkanes' resistance to flow, and one of their most physical properties. Understanding its connection with the alkanes' structure would enable manufacturing of superior oil and gas products, such as lubricants. However, alkanes are most often found in mixtures, and difficulties in separating their isomers with more than 12 carbon atoms prevents experimental measurements of their viscosity. Moreover, as no theoretical method is capable of predicting alkanes' viscosity, several semi-empirical methods have been developed to model it [18] [19] [68] [75] [99]. Unfortunately, the utility of these methods is still dependent on the quantity of experimental data, so they cannot be used to extrapolate to heavy alkanes.

Consequently, molecular dynamics simulations, in which the empirically parametrized force fields govern the Newtonian equations of motion for atoms and molecules, are used to model the real life behaviour of alkanes and model their viscosity. Molecular dynamics can be performed in equilibrium (EMD)[76] and out of equilibrium (NEMD)[92], and both approaches have been used to model viscosity of a range of alkanes [5] [27] [32] [64] [93] [94] [95] [97] [101] [33]. However, as the computational cost of the EMD grows quickly with the increase in molecular weight [76], EMD is used mostly to model viscosity of light alkanes.

The same, however, is not the case for the NEMD, which has been applied to heavy alkanes. For example, Cui et al. modelled dynamic viscosity of tetracosane and of 10-n-hexylnonadecane at 60°C, and of squalane at 38°C and at 99°C [29] [28]; Khare et al. of n-docosane, n-octacosane, and 5,12-dipropylhexadecane [63]; Scott, McCabe, and Cummings of squalane[11]; Lahtela et al. of eicosane isomers [69]; Moore et al. of n-triacontane, 9-octyldocosane, and squalane[90], and Kioupis et al. of 3-poly-α-olefin isomers as a function of pressure and temperature, [66] [65]; Liu et al. determined a pressure-viscosity coefficient of a 1-decene trimer [73], while Yang et al. determined a viscosity index of 2,2,4,4,6,8,8-heptamethylnonane[128].

Unfortunately, the predictive limits of NEMD viscosity simulations are still unknown, as NEMD viscosity results either have no experimental data available for comparison, or the simulations have not been performed independent of experiments, which allows to perform the simulations at as many manually chosen shear rates as possible to obtain better agreement with the experiments. In this chapter, however, we look to test the predictive limits of NEMD for the SciPCFF force field (subsection 1.4.2) by performing viscosity simulations for 11-heptyltricosane, 8,11-dipentyloctadecane, and 8,14-dipentylhenicosane independent of experiments, which have been performed by the BP scientists in Hull. In section 4.2, we

describe the calculation of viscosity state points, in section 4.3 we present the simulation results, while in section 4.4 we offer concluding remarks.

4.2 Preparing NEMD simulations

To obtain viscosity simulation state points, we first follow the procedure from subsection 1.4.3 to perform the molecular dynamics density simulations for 141 linear, single-branched, and double-branched alkanes at 25°C and at 100°C. Although some alkanes modeled are likely solids at 25°C, density of alkanes is a continuous function of temperature in liquid and solid phases, which justifies the use of molecular dynamics simulations even if we are performing simulations in a wrong phase.

Instead of performing molecular dynamics density simulations for 11-heptyltricosane, 8,11-dipentyloctadecane, and 8,14-dipentylhenicosane, we predict their molecular dynamics density values with neural networks (section 1.3) from molecular basis (subsection 2.1.3) to assess whether they can be used to prepare viscosity simulation state points without needing to perform density simulations every time, which would simplify and speed up the overall simulation process. The neural network models have previously been trained on molecular dynamics data for 141 alkanes, and its optimal structure with 6 hidden nodes found with a 5-fold cross-validation (subsection 1.3.3, [57]).

Neural network models accurately reproduce the simulation data, with the coefficient of determination R^2=0.996 at both temperatures, the average absolute deviation of is 1.16 $\frac{g}{l}$ at 25°C and of 1.21 $\frac{g}{l}$ at 100°C, with their standard deviations of 0.86$\frac{g}{l}$ and 1.02$\frac{g}{l}$, meaning that a lot of time can be saved in preparing the viscosity simulations when we use the SciPCFF force field. These results also demonstrate that molecular dynamics simulations exhibit regularity, verifying that the simulations have been performed properly, and display that molecular basis (subsection 2.1.3) efficiently encodes molecular structure for the purposes of density predictions.

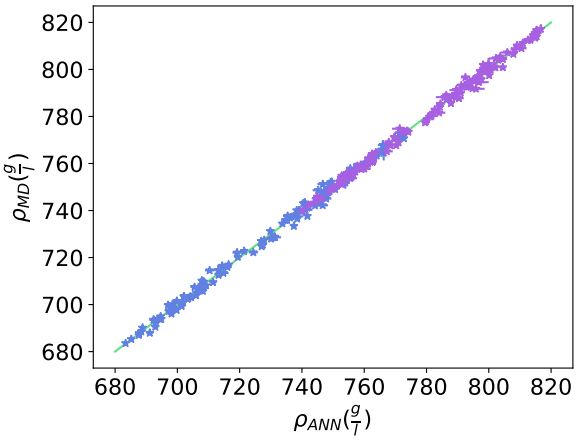


Fig. 4.1 Molecular dynamics density results plotted against the 5-fold cross-validation neural network results. Data entries at 25°C are highlighted in blue, while the ones at 100°C are highlighted in purple.

4.3 Results

We can now perform NEMD viscosity simulations for 11-heptyltricosane, 8,11-dipentyloctadecane, and 8,14-dipentylhenicosane. To perform the simulations, we follow the procedure from subsection 1.4.5 and use the sampling algorithm to select the good shear rates (subsection 3.3.1), while the density state points are given by either the mean neural network density prediction at 100°C or the linear interpolation of densities at 25°C and at 100°C. The linear interpolation of density is justified by the empirically observed linear dependence of alkanes' density on temperature.

After performing the simulations, we first note that the maximum total simulation time and the number of simulations are 46ns and 11 for the viscosity of 8,14-dipentylhenicosane at 40°C, as compared to the simulation time of 7ns and the total number of simulations of 7 required to simulate viscosity of hexane at 20°C (chapter 3), the smallest molecule considered in our study. This further shows why NEMD is the currently preferred method to simulate high viscosities. Simulation results are also in an excellent agreement with the Carreau model (Equation 1.22), with the coefficient of determination $R^2 \geq 0.99$ for all the simulations performed. A relatively small increase in the number of required simulations and the total physical simulation time when simulating viscosity of heavy rather compared to light alkanes, as well as an excellent agreement of simulation results with the Carreau model are an additional mark of the merit of the sampling algorithm (section 3.3).

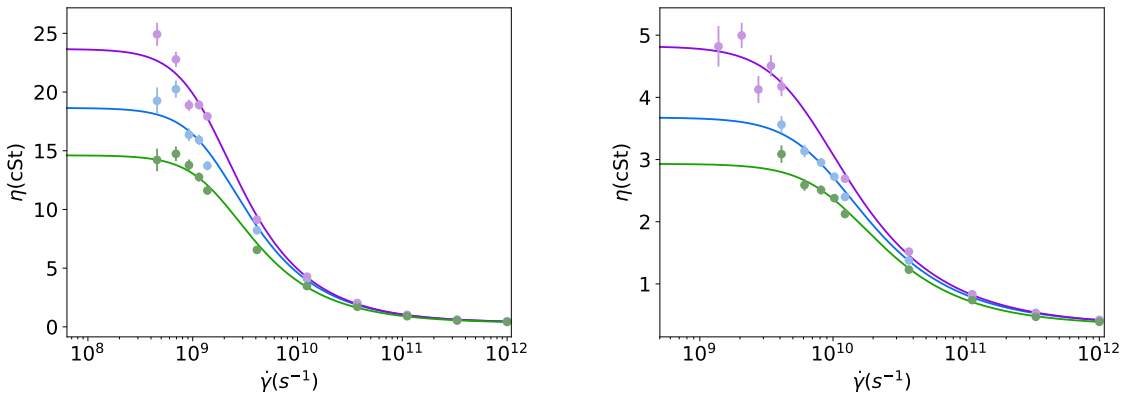


Fig. 4.2 Viscosity of 11-heptyltricosane (green), 8,11-dipentyloctadecane (blue), and 8,14-dipentylhenicosane (violet), and their best Carreau fit at 40°C(left) and 100°C(right). Shear rates are plotted on the logarithmic scale.

While the simulation results signal that branching affects the viscosity more heavily than molecular weight (Figure 4.2), the ultimate test of molecular dynamics comes from comparing its predictions to experimental data. Simulations are in good agreement with the experiments for 11-heptyltricosane, with percent errors of 8.8% at 40°C and of 16% at 100 °C, but are less accurate for 8,11-dipentyloctadecane and 8,14-dipentylhenicosane, with percent errors of 55% at 40°C and of 23% at 100°C for 8,11-dipentyloctadecane and of 39% at 40°C and 28% at 100°C for 8,14-dipentylhenicosane (Table 4.1).

We cannot determine the exact contributions of distinct sources to discrepancy between the simulations and the experiments, but it is highly likely that the systematic error in the SciPCFF force field is its main source. Since the accuracy of molecular dynamics density simulations is unknown, we do not know whether the simulations have been performed at correct external pressures (subsection 3.3.2). Also, the ranking of molecules' simulation viscosity values is constant as a function of the shear rates at both temperatures, so bad choice of shear rates cannot explain most of the discrepancies either.

Name	T(°C)	ρ ($\frac{g}{l}$)	η_{exp} (cSt))	η_{sim} (cSt)	$\delta\eta_{sim}$ (cSt)	$\Delta_{\%}$
8,11-dipentyloctadecane	40	792.29	12.02	18.65	1.38	55
8,11-dipentyloctadecane	100	753.47	2.99	3.68	0.28	23
8,14-dipentylhenicosane	40	794.4	16.97	23.66	1.13	39
8,14-dipentylhenicosane	100	756.28	3.77	4.83	0.22	28
11-heptyltricosane	40	787.9	13.46	14.61	1.11	8.8
11-heptyltricosane	100	750.58	3.47	2.93	0.21	-16

Table 4.1 Summary of viscosity results

From viscosities at 40°C and 100°C, we can predict the alkanes' viscosity index (VI)

$$VI = 100\frac{L-U}{L-H} \tag{4.1}$$

with U the kinematic viscosity at 40°C, and L and H obtained from the kinematic viscosity at 100°C [25]. To determine σ_{VI}, the uncertainty in the viscosity index, we propagate the uncertainties in L, U, and H

$$\sigma_{VI} = VI\sqrt{\frac{\sigma_L^2+\sigma_U^2}{(L-U)^2}+\frac{\sigma_H^2+\sigma_L^2}{(L-H)^2}-2\frac{\sigma_L^2-\sigma_{LH}^2}{(L-U)(L-H)}} \tag{4.2}$$

where we have used bilinearity of the covariance function, as well as Cov$[L,H]$=Cov$[L,U]$=0. Expected values and the elements of the L and H covariance matrix are calculated by taking 100 random samples from the normal distribution of kinematic viscosity at 100°C. After comparing our predictions with the experiments, we observe that the viscosity index of 8,14-dipentylhenicosane is the most accurate, with the 12% error and the experimental viscosity index within the predicted statistical uncertainty, despite the low accuracy of 8,14-dipentylhenicosane viscosity simulation results, since the discrepancies partially cancel. Unfortunately, the same is not the case for the other two molecules, and in particular for 11-heptyltricosane, for which the different signs of discrepancies at two temperatures have resulted in the viscosity index that is of a different order of magnitude to its experimental value (Table 4.2).

Name	VI_{exp}	VI_{model}
8,11-dipentyloctadecane	85	65±57
8,14-dipentylhenicosane	115	129±19
11-heptyltricosane	145	12±64

Table 4.2 Summary of viscosity index results

4.4 Conclusion

The results in this short chapter illuminated the current applicability of neural networks and molecular dynamics in predicting the kinematic viscosity of large alkanes. Neural networks have shown that they can reproduce molecular dynamics density results, and the sampling algorithm has shown consistency, speed and reliability. These are encouraging signs for the future of NEMD. Unfortunately, the viscosity simulation results still do not agree with the experiments due to a poor force field. To resolve this issue, one can collect more experimental density and viscosity data and exploit the patterns in discrepancy between simulations and experiments, like in section 3.2. However, a probably better approach would be to develop more accurate force fields, through either machine learning [122] or DFT [55] [84]. In any case, the accuracy of force fields remains a major obstacle that needs to be resolved before NEMD viscosity simulations can be used as an accurate substitute for experiments.

Chapter 5

Theory of uniform mixtures of molecular liquids

We develop a predictive theory of intermolecular forces dependent physical properties of mixtures whose molecules are uniformly spatially distributed. To test its validity, we apply this theory to mixtures of alkanes by fitting it to experimental mixture data for molar volume, surface tension, isentropic compressibility, and dynamic viscosity, and then using the best fit parameters for dynamic viscosity to predict dynamic viscosity of other mixtures and compare to experiments. Theoretical predictions and best fits show excellent agreement with experiments, with the average absolute percent deviation equal to at most 0.8% for predictions and to 0.2% for fits. These results suggest that our theory can be used to accurately model and predict many physical properties of mixtures of alkanes, while the conceptual basis behind it holds promise to be applicable to more complex mixtures.

5.1 Introduction

Liquid mixtures are ubiquitous in nature. From the blood that pumps through our veins to a cup of coffee or tea that give us a head start every morning, they play an essential role in our lives. In addition to their everyday importance, they are of great interest to scientists, as the interactions between unlike substances give rise to property values that are different to the weighted average of pure component properties [87].

Molecular liquids comprise a large subset of liquid mixtures, yet many of their property values are unknown. Difficulties in separating them prevents empirical analysis of their properties. Moreover, while several theoretical formalisms have been developed to model mixture properties [100] [86] [110] [44] [62], most commonly molar volume and dynamic viscosity of binary alkane mixtures, their predictive power is limited, and none of them apart from the Flory theory [100] deliver insights into the underlying physics.

A canonical insight into the mixture physics is the effect of spatial probability distribution of individual molecules on their physical properties that depend on intermolecular forces. In this chapter, we use this insight to develop a theory of mixtures whose individual molecules follow a uniform spatial probability distribution, called uniform mixtures. Then, we apply this theory to mixtures of alkanes, and describe ideas for its future applications and extensions.

5.2 Theory

Physical properties of different phases of matter are usually described in terms of interactions between the matters' basic constituents, such as atoms or molecules. Two examples of physical systems at the extremes of interactions between the particles are the ideal gas, where the widely separated molecules do not interact, and the crystalline solid, where densely packed atoms form equally spaced repeating patterns and strongly interact. Molecular liquids lie between these two extremes, with their molecules interacting, but irregularly due to uneven spacing between the molecules. These irregularities mean that, unlike for the ideal gas or the crystalline solid, there is at present no theory of physical properties of molecular liquids, so the best we can do is to express the ones that depend on intermolecular forces [56] as an expectation value

$$\mathbb{E}[Q] = \mathbb{E}[f_Q(\vec{F}_{II}(\vec{\alpha})], \tag{5.1}$$

where Q represents a physical property, $\mathbb{E}$ the expectation operator, and f_Q an unknown property specific function of the intermolecular forces $\vec{F}_{II}(\vec{\alpha})$ between the same molecular species, described by a set of parameters $\vec{\alpha}$. The expectation value is taken over all possible

relative molecular orientations and distances, and the variance in Q is neglected, since the system is macroscopic and does not exhibit long range orientational order.

To obtain a similar expression for the physical properties of uniform mixtures of molecular liquids, we initially look into the local property contributions due to interactions with species I. Working under a physically realistic assumption that the effect of intermolecular forces is negligible beyond a cutoff radius, and considering only the 2-body interactions, its average Q_I is given by

$$Q_I = \frac{1}{\tilde{N}-1}\left[\sum_{i=1}^{\tilde{N}_I-1} f_Q(\vec{F}_{II}(\vec{\alpha}_i)) + \sum_{J\neq I}\sum_{i=1}^{\tilde{N}_J} f_Q(\vec{F}_{IJ}(\vec{\alpha}_i))\right], \tag{5.2}$$

such that $\tilde{N}$ symbolises the total number of molecules within the cutoff radius, $\tilde{N}_I$ and $\tilde{N}_J$ the number of molecules of each type within the cutoff radius, while $f_Q(\vec{F}_{II}(\vec{\alpha}))$ and $f_Q(\vec{F}_{IJ}(\vec{\alpha}))$ represent the function of forces between the molecular species I and J. We determine the expectation value of this contribution by conditioning over f_Q and $\tilde{N}$. Under the assumption that the spatial probability distribution of each molecule is uniform, and with the conservation of expectation value of the function of intermolecular forces between the like molecules upon mixing, we obtain

$$\mathbb{E}[Q_I] = \frac{1}{N-1}\left[(N_I-1)\,\mathbb{E}[Q_{II}] + \sum_{I\neq J} N_J\,\mathbb{E}[Q_{IJ}]\right], \tag{5.3}$$

where $\mathbb{E}[Q_{II}]$ is the property value of I species (Equation 5.1), $\mathbb{E}[Q_{IJ}] = \mathbb{E}[f_Q(\vec{F}_{IJ}(\vec{\alpha}))]$ is attributed to interactions between unlike molecules, while N, N_I, N_J are the total number of molecules, and the number of molecules of each type. We can now obtain the expression for the physical properties of uniform mixtures. By averaging over all the local contributions ($Q = \sum_I \frac{N_I}{N} Q_I$), and with $Q_{JI} = Q_{IJ}$, we find that

$$\mathbb{E}[Q] = \sum_I x_I^2\,\mathbb{E}[Q_{II}] + 2\sum_{I<J} x_I x_J\,\mathbb{E}[Q_{IJ}], \tag{5.4}$$

where $\{x_I = \frac{N_I}{N}\}$ denote the molar fraction of molecular species. Since $\{\mathbb{E}[Q_{IJ}]\}$ depend only on molecular species, theory of uniform mixtures of molecular liquids enables prediction of multicomponent mixture properties from the binary mixture properties, and vice versa. Additionally, two familiar expressions arise as special cases of Equation 5.4, with the Raoult's law of ideal mixtures appearing when $Q_{IJ} = \frac{Q_{II}+Q_{JJ}}{2}\ \forall\ I\neq J$, and the empirical expression from Hind et al. [50] arising when Equation 5.4 models dynamic viscosity of binary mixtures.

5.3 Application to mixtures of alkanes

We now apply the theory of uniform mixtures of molecular liquids to mixtures of alkanes. These mixtures are miscible and non-polar, with the London dispersion force the only intermolecular interaction, allowing us to approximate the spatial distribution of their molecules as uniform. Also, they have a considerable amount of available experimental data relative to other types of mixtures, which enables a more in depth comparison of theory to experiment.

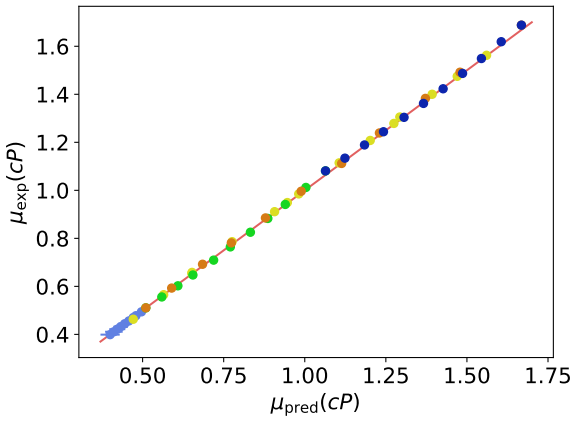
To assess the accuracy of the theory of uniform mixtures of molecular liquids, we fit Equation 5.4 to experimental molar volume [10], surface tension [34], isentropic compressibility [42], and dynamic viscosity [126] [22] data with linear regression [57] without the bias parameter, molar fraction terms as independent variables, and $\{\mathbb{E}[Q_{IJ}]\}$ as model parameters. The best fits show an excellent agreement with the experiments, with the maximum average absolute percent error $\Delta_{|\%|}$ (section 3.1) of 0.2% for the viscosity of the mixture of heptane, octane, undecane, and tridecane (Appendix D). Also, the pure component properties obtained as best fit parameters are within experimental uncertainties for the surface tension and isentropic compressibility, and still in a very good agreement for the molar volume and dynamic viscosity, with the largest $\Delta_{|\%|}$ of 1.5% for viscosity of undecane (Table 5.1). Then, as a simple assessment of overfitting (subsection 1.3.3), we fit Equation 5.4 to mixtures that have a pure component in common. Since the best fit pure component properties are independent of the mixture used to find them, with the worst agreement of 1.9% for hexane and of at most 0.2% for other pure component properties (Table 5.1), we deduce that Equation 5.4 did not overfit the experimental data.

Property	**Mixture**	T(°C)	$\mathbb{E}[Q_{II}]$	$\mathbb{E}[Q_{IJ}]$
$V_{\mathrm{m}}(10^{-6}\mathrm{m^3mol^{-1}})$	nonane isooctane	25	179.62±0.01 (179.61±0.01) 166.13±0.01* (166.14±0.01)	172.57±0.01
$V_{\mathrm{m}}(10^{-6}\mathrm{m^3mol^{-1}})$	dodecane isooctane	25	228.60±0.01 (228.57±0.01) 166.10±0.01* (166.14±0.01)	196.67±0.02
$\sigma(10^{-3}\mathrm{kg\ s^{-2}})$	heptane isooctane	40	18.16 (18.15±0.03) 17.23‡ (17.23±0.03)	17.62±0.01
$\sigma(10^{-3}\mathrm{kg\ s^{-2}})$	octane isooctane	40	19.82 (19.82±0.03) 17.23‡ (17.23±0.03)	18.44
$K_{\mathrm{S,m}}$ $(10^{-6}\mathrm{kg\ m^3mol^{-1}s^{-2}})$	nonane toluene	10	151720±40 (151700±300) 63850±30| (63910±130)	109590±30
$K_{\mathrm{S,m}}$ $(10^{-6}\mathrm{kg\ m^3mol^{-1}s^{-2}})$	hexane toluene	10	147000±60 (147000±290) 63900±50| (63910±130)	95710±100

$\mu(10^{-3}\text{kg m}^{-1}\text{ s}^{-1})$	heptane octane undecane tridecane	25	0.3886±0.0323 (0.3889±0.0004) 0.5099±0.0077 (0.5105±0.0005) 1.0648±0.0063 (1.0810±0.0011) 1.6684±0.0079 (1.6880±0.0017)	μ_{7+8}=0.4411±0.0201 μ_{7+11}=0.6224±0.0181 μ_{7+13}=0.7562±0.0163 μ_{8+11}=0.7657±0.0145 μ_{8+13}=0.8826±0.0147 μ_{11+13}=1.3636±0.0132
$\mu(10^{-3}\text{kg m}^{-1}\text{ s}^{-1})$	hexane tetradecane	25	0.3198[§] (0.2920±0.0003) 2.0460[¶] (2.0620±0.0021)	0.6975
$\mu(10^{-3}\text{kg m}^{-1}\text{ s}^{-1})$	hexane hexadecane	25	0.3258[§] (0.2920±0.0003) 3.0159[**] (3.0480±0.0030)	0.7964
$\mu(10^{-3}\text{kg m}^{-1}\text{ s}^{-1})$	tetradecane hexadecane	25	2.0445[¶] (2.0620±0.0021) 3.0209[**] (3.0480±0.0030)	2.4857

Table 5.1 Summary of fitting parameters of Equation 5.4 to experiments for the molar volume (V_m), surface tension (σ), molar isentropic compressibility ($K_{\text{S,m}}$), and dynamic viscosity (μ). Q_{II} represent property values of pure components and Q_{IJ} the terms arising from intermolecular interactions between unlike molecules. Experimental values of pure component properties are found in the parentheses. Uncertainties in the best fit parameters for viscosity of binary mixtures and surface tension are zero to four, and two decimal places, so are not reported. If the pure property value of a pure liquid has been obtained twice by fitting Equation 5.4 to experimental data of multiple mixtures, it is highlighted with a bespoke superscript.

With the fitting accuracy of Equation 5.4 established, we now use the best fit parameters for viscosity Table 5.1 to assess our theory's predictive power. Specifically, we use the best fit parameters from the hexane+tetradecane, hexane+hexadecane, and tetradecane+hexadecane mixtures to predict the dynamic viscosity of the hexane+tetradecane+hexadecane mixture at 25°C, and the best fit parameters from the heptane+octane+undecane+tridecane mixture to predict the dynamic viscosity of six binary (heptane+octane, heptane+undecane, heptane+tridecane, octane+undecane, octane+tridecane, undecane+tridecane) and four ternary (heptane+octane+undecane, heptane+octane+tridecane, heptane+undecane+tridecane, octane+undecane+tridecane) mixtures at the same temperature. For the hexane+tetradecane+hexadecane, heptane+octane, octane+undecane, octane+tridecane, and undecane+tridecane mixtures, which have available experimental data, we compare predictions to experiments, and note that they are in excellent agreement, with $\Delta_{|\%|} \leq 0.8\%$ for all of them (Figure 5.1).



Mixture	$\Delta_{\|\%\|}$
Hexane, tetradecane, hexadecane	0.6
Heptane, octane	0.2
Octane, undecane	0.8
Octane, tridecane	0.7
Undecane, tridecane	0.3

Fig. 5.1 Experimental values plotted against theoretical predictions of viscosity of mixtures of hexane, tetradecane, and hexadecane (yellow), heptane and octane (light blue), octane and undecane (green), octane and tridecane (orange), and undecane and tridecane (dark blue) are shown on the left, while the statistical summary of predictions is show on the right. The parity line $\mu_{exp} = \mu_{pred}$ is also plotted as a reference, to help visualize the accuracy of predictions.

5.4 Conclusion

In this chapter, we have developed a theory of uniform mixtures of molecular liquids, and applied it to mixtures of alkanes. The theory was fitted to experimental data for the molar volume, surface tension, isentropic compressibility, and dynamic viscosity, and the best fit viscosity parameters were used to predict the viscosity of other mixtures. Predictions and theoretical fits are in an excellent agreement with the experiments, showing that our theory is an excellent description of many physical properties of mixtures of alkanes when the pure components are far away from their phase transition points.

Work presented in this chapter offers new insights and has the potential to open new avenues in mixture research. From the applied perspective, experimentally measuring properties of additional mixtures of alkanes would allow us to test the theory of uniform mixtures further, as well as to investigate the effects of mixture composition on the interaction constants between unlike molecules. The latter would allow us to extrapolate the interaction constants between unlike molecules with machine learning section 1.3, and predict many properties of alkane mixtures far away from the phase transition points of pure alkanes. More importantly, work in this chapter suggests that mixtures' deviation from ideal behaviour is driven by the spatial probability distribution of their molecules. With this in mind, different probability distributions can be used to develop theories of different types of mixtures, and

tested against experiments. If these theories are proven correct, they would illuminate the currently unexplored rich world of mixtures' physics.

Chapter 6

Conclusion

In this Thesis, we have applied neural networks and molecular dynamics to model the physical properties of pure alkanes (chapter 2, chapter 3, and chapter 4). We have also developed a theory of mixtures of molecular liquids whose molecules' positions follow a uniform probability distribution, and applied it to mixtures of alkanes in chapter 5. The approach to model properties of molecular systems with a cocktail of computational methods is likely to grow in popularity and importance in the upcoming years, especially as the demand for new materials increases. Still, even if our results are encouraging, the lack of experimental data and low accuracy of molecular dynamics force fields still remain major obstacles. The theory of uniform mixtures (chapter 5) can be applied to further mixtures of alkanes, subject to availability of experimental data, while the conceptual basis behind it can be further exploited to construct theories of more complex mixtures. However, the experimental data is currently scarse, and theories of more complex mixtures will likely require significant effort.

Overall, we should be content with the progress made in understanding the physical properties of alkanes and their mixtures. However, we should also be aware that the full implications of the work presented in this Thesis are unlikely to be immediately known, while our work will not instantly lead to production of superior lubricants. This is neither emotionally nor economically satisfying. Unfortunately, scientific progress is neither linear nor instant, contrary to how it is viewed by the majority of people. Many discoveries are unplanned, and sometimes even accidental. However, if the modern society genuinely wishes to fully exploit the potential of scientific knowledge, we should first all accept that research is often arcane and confusing, even for the experts, that there are many phenomena we still cannot explain or even know about, and that we cannot always force discoveries, no matter how much we want to.

References

[1] Amber force field. "https://ambermd.org/".

[2] CHARMM froce field. https://www.charmm.org/".

[3] GROMOS force field. "http://www.gromos.net/".

[4] American Petroleum Institute. Research Project 44. *TRC Thermodynamic Tables: Hydrocarbons*. Thermodynamics Research Center, Texas Engineering Experiment Station, Texas A & M University System, 1986.

[5] W. Allen and R.L. Rowley. Predicting the viscosity of alkanes using nonequilibrium molecular dynamics: Evaluation of intermolecular potential models. *The Journal of Chemical Physics*, 106(24):10273–10281, 1997.

[6] S.S. Alqaheem and M.R. Riazi. Flash points of hydrocarbons and petroleum products: Prediction and evaluation of methods. *Energy & Fuels*, 31(4):3578–3584, 2017.

[7] A. Arora. *Hydrocarbons: Alkanes, Alkenes and Alkynes*. Discovery Publishing House, New Delhi, 2006.

[8] M.J. Assael and M. Papadaki. Measurements of the viscosity of n-heptane, n-nonane, and n-undecane at pressures up to 70 MPa. *International Journal of Thermophysics*, 12(5):801–810, 1991.

[9] P. W. Atkins and J. De Paula. *Atkins' Physical Chemistry (8th ed.)*. W.H. Freeman, New York, 2006.

[10] A.M. Awwad and E.I. Allos. Thermodynamic properties of binary mixtures of isooctane and n-alkane at 298.15 K. *Fluid Phase Equilibria*, 22(3):353–366, 1985.

[11] S. Bair, C. McCabe, and P.T. Cummings. Comparison of nonequilibrium molecular dynamics with experimental measurements in the nonlinear shear-thinning regime. *Phys. Rev. Lett.*, 88(5), 2002.

[12] H.O. Baled, D. Xing, H. Katz, D. Tapriyal, I.K. Gamwo, Y. Soong, B.A. Bamgbade, Y. Wu, K. Liu, M.A. McHugh, and R.M. Enick. Viscosity of n-hexadecane, n-octadecane and n-eicosane at pressures up to 243MPa and temperatures up to 534K. *The Journal of Chemical Thermodynamics*, 72:108–116, 2014.

[13] G.K. Batchelor. *An Introduction to Fluid Dynamics*. Cambridge University Press, Cambridge, 2000.

[14] C.M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[15] K.J. Burch and E.G. Whitehead. Melting-point models of alkanes. *Journal of Chemical & Engineering Data*, 49(4):858–863, 2004.

[16] K.J. Burch, D.K. Wakefield, and E.G. Whitehead Jr. Boiling point models of alkanes. *MATCH Communications in Mathematical and in Computer Chemistry*, 1:25–52, 2003.

[17] J.G. Calvert, R.G. Derwent, J.J. Orlando, T.J. Wallington, and G.S. Tyndall. *Mechanisms of Atmospheric Oxidation of the Alkanes*. Oxford University Press, Oxford, 2008.

[18] D.R. Caudwell, J.P.M. Trusler, V.Vesovic, and W.A. Wakeham. The viscosity and density of n-dodecane and n-octadecane at pressures up to 200 MPa and temperatures up to 473 K. *International Journal of Thermophysics*, 25(5):1339–1352, 2004.

[19] D.R. Caudwell, J.P.M. Trusler, V. Vesovic, and W.A. Wakeham. Viscosity and density of five hydrocarbon liquids at pressures up to 200 MPa and temperatures up to 473 K. *Journal of Chemical & Engineering Data*, 54(2):359–366, 2009.

[20] D.R. Caudwell, J.P.M. Trusler, V. Vesovic, and W.A. Wakeham. Viscosity and density of five hydrocarbon liquids at pressures up to 200 MPa and temperatures up to 473 K. *Journal of Chemical & Engineering Data*, 54(2):359–366, 2009.

[21] R. Chang and K. Goldsby. *General Chemistry: The Essential Concepts (7th Edition)*. McGraw-Hill Education, New York, 2013.

[22] J.L.E. Chevalier, P.J. Petrino, and Y.H. Gaston-Bonhomme. Viscosity and density of some aliphatic, cyclic, and aromatic hydrocarbons binary liquid mixtures. *Journal of Chemical & Engineering Data*, 35(2):206–212, 1990.

[23] S. Cho, S. Jeong, J.M. Kim, and C. Baig. Molecular dynamics for linear polymer melts in bulk and confined systems under shear flow. *Scientific Reports*, 7(1)(9004), 2017.

[24] A. Choromanska, M. Henaff, M. Mathieu, G.B. Arous, and Y. LeCun. The loss surfaces of multilayer networks. https://arxiv.org/abs/1412.0233.

[25] COMPASS. https://compass.astm.org/EDIT/html_annot.cgi?D2270+10(2016).

[26] L. Constantinou and R. Gani. New group contribution method for estimating properties of pure compounds. *AIChE Journal*, 40(10):1697–1710, 1994.

[27] S.T. Cui, P.T. Cummings, and H.D. Cochran. The calculation of the viscosity from the autocorrelation function using molecular and atomic stress tensors. *Molecular Physics*, 88(6):1657–1664, 1996.

[28] S.T. Cui, S.A. Gupta, P.T. Cummings, and H.D. Cochran. Molecular dynamics simulations of the rheology of normal decane, hexadecane, and tetracosane. *The Journal of Chemical Physics*, 105(3):1214–1220, 1996.

[29] S.T. Cui, P.T. Cummings, H.D. Cochran, J.D. Moore, and S.A. Gupta. Nonequilibrium molecular dynamics simulation of the rheology of linear and branched alkanes. *International Journal of Thermophysics*, 19(2):449–459, 1998.

[30] H.B. Curry. The method od steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics*, 2(3):258–261, 1944.

[31] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems*, 2:303–314, 1989.

[32] P.J. Daivis and D.J. Evans. Comparison of constant pressure and constant volume nonequilibrium simulations of sheared model decane. *The Journal of Chemical Physics*, 100(1):541–547, 1994.

[33] P. Daugé, A. Baylaucq, X. Canet, and C. Boned. High pressure viscosity and density measurements of the binary mixture tridecane + 2,2,4,4,6,8,8-heptamethylnonane. *High Pressure Research*, 18(1-6):291–296, 2000.

[34] J. de los S. López-Lázaro, G.A. Iglesias-Silva, A. Estrada-Baltazar, and J. Barajas-Fernández. Density and surface tension of binary mixtures of 2,2,4-trimethylpentane + n-heptane, 2,2,4-trimethylpentane + n-octane, ethyl acetate + benzene, and butanenitrile + benzene from (293.15 to 323.15) K. *Journal of Chemical & Engineering Data*, 60(6):1823–1834, 2015.

[35] D.Frenkel and B.Smit. *Understanding Molecular Simulation (Second Edition)*. Academic Press, San Diego, 2002.

[36] D.Mathieu. Inductive modeling of physico-chemical properties: Flash point of alkanes. *Journal of Hazardous Materials*, 179(1):1161–1164, 2010.

[37] D. J. Evans and G.P. Moriss. *Statistical Mechanics of Nonequilibrium Liquids*. Academic Press, Cambridge, MA, USA.

[38] D.J. Evans and G.P. Morriss. Nonlinear-response theory for steady planar Couette flow. *Phys. Rev. A*, 30:1528–1530, 1984.

[39] W.A. Felsing and G.M. Watson. The compressibility of liquid n-octane. *Journal of the American Chemical Society*, 64(8):1822–1823, 1942.

[40] H. Flyvbjerg and H.G. Petersen. Error estimates on averages of correlated data. *The Journal of Chemical Physics*, 91(1):461–466, 1989.

[41] Y. Goldberg. *Neural Network Methods in Natural Language Processing*. Morgan & Claypool Publishers, San Rafael, California, 2017.

[42] B. González, E.J. González, I. Domínguez, and A. Domínguez. Excess properties of binary mixtures hexane, heptane, octane and nonane with benzene, toluene and ethylbenzene at t = 283.15K and 298.15 K. *Physics and Chemistry of Liquids*, 48(4): 514–533, 2010.

[43] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[44] L. Grunberg and A. Nissan. Mixture law for viscosity. *Nature*, 164:799–800, 1949.

[45] K. He, X. Zhang, S. Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.

[46] D.W. Heermann and K. Binder. *Monte Carlo Methods in Statistical Physics (6th edition)*. Springer-Verlag, Berlin, 2019.

[47] E.L. Heric and J.G. Brewer. Viscosity of some ternary liquid nonelectrolyte mixtures. *Journal of Chemical & Engineering Data*, 14(1):55–63, 1969.

[48] M.A. Hernández-Galván, F. García-Sánchez, and R. Macías-Salinas. Liquid viscosities of benzene, n-tetradecane, and benzene+n-tetradecane from 313K to 393K and pressures up to 60MPa: Experiment and modeling. *Fluid Phase Equilibria*, 261(1): 51–60, 2007.

[49] M.R. Hestenes and E.Stiefel. Methods of conjugate gradients for solving linear systems. 49(6), 1952.

[50] R.K. Hind, E.McLaughlin, and A.R. Ubbelohde. Structure and viscosity of liquids. viscosity-temperature relationships of pyrrole and pyrrolidine. *Trans. Faraday Soc.*, 56:331–334, 1960.

[51] R.W. Hockney and J.W. Eastwood. *Computer simulation using particles*. Taylor & Francis Group, New York, 1988.

[52] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251 – 257, 1991.

[53] A. Hospital, J.R. Goni, M. Orozco, and J.L. Gelpi. Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry*, 8:37–47, 2019.

[54] S.M. Hosseini, M. Pierantozzi, and J. Moghadasi. Viscosities of some fatty acid esters and biodiesel fuels from a rough hard-sphere-chain model and artificial neural network. *Fuel*, 235:1083–1091, 2019.

[55] R. Iftimie, P. Minary, and M.E. Tuckerman. Ab initio molecular dynamics: Concepts, recent developments, and future trends. *Proceedings of the National Academy of Sciences*, 102(19):6654–6659, 2005.

[56] J.N. Israelachvili. *Intermolecular and Surface Forces (Third Edition)*. Academic Press, Boston, 2011.

[57] J.Friedman, T.Hastie, and R.Tibshirani. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2009.

[58] W.L. Jorgensen and J. Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[59] W.L. Jorgensen and J. Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[60] J.Wei. Molecular symmetry, rotational entropy, and elevated melting points. *Industrial & Engineering Chemistry Research*, 38(12):5019–5027, 1999.

[61] J.Wu, A.H. Nhaesi, and A.F.A. Asfour. Viscosities of eight binary liquid n-alkane systems at 293.15K and 298.15K. *Journal of Chemical & Engineering Data*, 44(5): 990–993, 1999.

[62] P.K. Katti and M.M. Chaudhri. Viscosities of binary mixtures of benzyl acetate with dioxane, aniline and m-cresol. *J. Chem. Eng. Data*, 9:442–443, 1964.

[63] R. Khare, J. De Pablo, and A. Yethiraj. Rheological, thermodynamic, and structural studies of linear and branched alkanes under shear. *The Journal of Chemical Physics*, 107(17):6956–6964, 1997.

[64] L.I. Kioupis and E.J. Maginn. Rheology, dynamics, and structure of hydrocarbon blends: a molecular dynamics study of n-hexane/n-hexadecane mixtures. *Chemical Engineering Journal*, 74:129–146, 1999.

[65] L.I. Kioupis and E.J. Maginn. Molecular simulation of poly-α-olefin synthetic lubricants: Impact of molecular architecture on performance properties. *The Journal of Physical Chemistry B*, 103:10781–10790, 1999.

[66] L.I. Kioupis and E.J. Maginn. Impact of molecular architecture on the high-pressure rheology of hydrocarbon fluids. *The Journal of Physical Chemistry B*, 104:7774–7783, 2000.

[67] N.D. Kondratyuk, A.V. Lankin, G.E. Norman, and V.V. Stegailov. Relaxation and transport properties of liquid n-triacontane. *Journal of Physics: Conference Series*, 653(012107), 2015.

[68] J.J. De la Porte and C.A. Kossack. A liquid phase viscosity–temperature model for long-chain n-alkanes up to C64H130 based on the free volume theory. *Fuel*, 135: 156–164, 2014.

[69] M. Lahtela, M. Linnolahti, T.A. Pakkanen, and R.L. Rowley. Computer simulations of branched alkanes: The effect of side chain and its position on rheological behavior. *The Journal of Chemical Physics*, 106(6):2626–2630, 1998.

[70] A.W. Lees and S.F. Edwards. The computer study of transport processes under extreme conditions. *J. Phys. C: Solid State Phys.*, 5(1921), 1972.

[71] K. Leveberg. A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math.*, 2:164–168, 1944.

[72] D.C. Liu and J.Nocedal. On the limited memory bfgs method for large scale optimization. 45:503–528, 1989.

[73] P. Liu, H. Yu, R. Ning, F.E. Lockwood, and J.Q. Wang. Pressure-viscosity coefficient of hydrocarbon base oil through molecular dynamics simulations. *Tribology Letters*, 60(3)(34), 2015.

[74] P. Liu, H. Yu, N.Ren, F.E. Lockwood, and J.Q. Wang. Lubricant shear thinning behavior correlated with variation of radius of gyration via molecular dynamics simulations. *The Journal of Chemical Physics*, 147(8)(084904), 2017.

[75] L.T.Novak. Predictive corresponding-states viscosity model for the entire fluid region: n-alkanes. *Industrial & Engineering Chemistry Research*, 52(20):6841–6847, 2013.

[76] E.J. Maginn, R.A. Messerly, D.J. Carlson, D.R. Roe, and R.J. Elliott. Best practices for computing transport properties 1. Self-diffusivity and viscosity from equilibrium molecular dynamics. *Living Journal of Computational Molecular Science*, 1, 2018.

[77] J.J. Marano and G.D. Holder. General equation for correlating the thermophysical properties of n-paraffins, n-olefins, and other homologous series. 1. Formalism for developing asymptotic behavior correlations. *Industrial & Engineering Chemistry Research*, 36(5):1887–1894, 1997.

[78] J.J. Marano and G.D. Holder. General equation for correlating the thermophysical properties of n-paraffins, n-olefins, and other homologous series. 2. Asymptotic behavior correlations for pvt properties. *Industrial & Engineering Chemistry Research*, 36(5):1895–1907, 1997.

[79] J.J. Marano and G.D. Holder. A general equation for correlating the thermophysical properties of n-paraffins, n-olefins, and other homologous series. 3. Asymptotic behavior correlations for thermal and transport properties. *Industrial & Engineering Chemistry Research*, 36(6):2399–2408, 1997.

[80] D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[81] Sonsoles Martín-Santamaría. *Computational Tools for Chemical Biology*. Royal Society of Chemistry, London, 2017.

[82] L. M. Martins. *Alkanes: Properties, Production and Applications*. Nova Science Publishers, New York, 2019.

[83] G.J. Martyna, M.L. Klein, and M. Tuckerman. Nosé–hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics*, 97(4):2635–2643, 1992.

[84] D. Marx and J. Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, Cambridge,UK, 2009.

[85] M.A. Matthews, J.B. Rodden, and A. Akgerman. High-temperature diffusion, viscosity, and density measurements in n-hexadecane. *Industrial & Engineering Chemistry Research*, 32(3):317–319, 1987.

[86] R.A. McAllister. The viscosity of liquid mixtures. *AIChE J.*, 6:427–431, 1960.

[87] D.A. McQuarrie. *Physical chemistry : a molecular approach*. University Science Books, Sausalito, California, 1997.

[88] R.A. Messerly, T.A. Knotts, N.F. Giles, and W.V. Wilding. Developing an internally consistent set of theoretically based prediction models for the critical constants and normal boiling point of large n-alkanes. *Fluid Phase Equilibria*, 449:104–116, 2017.

[89] M.Leshno, V.Y. Lin, A. Pinkus, and S. Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861 – 867, 1993.

[90] J.D. Moore, S.T. Cui, H.D. Cochran, and P.T. Cummings. Rheology of lubricant basestocks: A molecular dynamics study of c30 isomers. *The Journal of Chemical Physics*, 113(19):8833–8840, 200.

[91] J.D. Moore, S.T. Cui, H.D. Cochran, and P.T. Cummings. A molecular dynamics study of a short-chain polyethylene melt.: I. Steady-state shear. *Journal of Non-Newtonian Fluid Mechanics*, 93(1):83–99, 2000.

[92] G.P. Morriss and D.J. Evans. A constraint algorithm for the computer simulation of complex molecular liquids. *Computer Physics Communications*, 62(2):267–278, 1991.

[93] C.J. Mundy, J.I. Siepmann, and M.L. Klein. Decane under shear: A molecular dynamics study using reversible NVT-SLLOD and NPT-SLLOD algorithms. *The Journal of Chemical Physics*, 103(23):10192–10200, 1995.

[94] C.J. Mundy, S. Balasubramanian, K. Bagchi, J.I. Siepmann, and M.L. Klein. Equilibrium and non-equilibrium simulation studies of fluid alkanes in bulk and at interfaces. *Faraday Discuss.*, 104(17), 1996.

[95] C.J. Mundy, M.L. Klein, and J.I. Siepmann. Determination of the pressure-viscosity coefficient of decane by molecular simulation. *J. Phys Chem*, 100(16779), 1996.

[96] D.E. Needham, I.C. Wei, and P.G. Seybold. Molecular modeling of the physical properties of alkanes. *Journal of the American Chemical Society*, 110(13):4186–4194, 1988.

[97] D. Nevins and F.J. Spere. Accurate computation of shear viscosity from equilibrium molecular dynamics simulations. *Molecular Simulation*, 33(15):1261–1266, 2007.

[98] M.P Nightingale and C.J Umrigar. *Quantum Monte Carlo Methods in Physics and Chemistry*. Springer, Dordrecht, 1999.

[99] N.Riesco and V.Vesovic. Extended hard-sphere model for predicting the viscosity of long-chain n-alkanes. *Fluid Phase Equilibria*, 425:385–392, 2016.

[100] R.A. Orwoll and P.J. Flory. Thermodynamic properties of binary mixtures of n-alkanes. *Journal of the American Chemical Society*, 89(26):6822–6829, 1967.

[101] G. Pan and C. McCabe. Prediction of viscosity for molecular fluids at experimentally accessible shear rates using the transient time correlation function formalism. *The Journal of Chemical Physics*, 125(19)(194527), 2006.

[102] R.G. Parr and W. Yang. *Density Functional Theory of Atoms and Molecules*. Oxford Science Publications, Oxford, 1989.

[103] R.K. Pathria and P.D. Beale. *Statistical Mechanics (Third Edition)*. Academic Press, Boston, 2011.

[104] R.S. Payal, S. Balasubramanian, I. Rudra, T. Kunj, I. Mahlke, D. Doyle, and R. Cracknell. Shear viscosity of linear alkanes through molecular simulations: quantitative tests for n-decane and n-hexadecane. *Molecular Simulation*, 38(14-15):1234–1241, 2012.

[105] Chemeo: Chemical & physical properties by chemeo. URL https://www.chemeo.com/. https://www.chemeo.com/.

[106] A. Pinkus. Approximation theory of the mlp model in neural networks.

[107] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117:1–19, 1995.

[108] A.J.L. Pombeiro and M.F.C.G. da Silva. *Alkane Functionalization*. John Wiley & Sons Ltd, New Jersey, 2018.

[109] The pubchem project. URL https://pubchem.ncbi.nlm.nih.gov/. https://pubchem.ncbi.nlm.nih.gov/.

[110] A.J. Queimada, S.E. Quiñones-Cisneros, I.M. Marrucho, J.A.P. Coutinho, and E.H. Stenby. Viscosity and liquid density of asymmetric hydrocarbon mixtures. *International Journal of Thermophysics*, 24:1221–1239, 2003.

[111] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. *Nature*, 323(6088):533–536, 1986.

[112] V. Ruzicka and E.S. Domalski. Estimation of the heat capacities of organic liquids as a function of temperature using group additivity. ii. Compounds of carbon, hydrogen, halogens, nitrogen, oxygen, and sulfur. *Journal of Physical and Chemical Reference Data*, 22(3):619–657, 1993.

[113] P. Santak and G. Conduit. Predicting physical properties of alkanes with neural networks. *Fluid Phase Equilibria*, 501(112259), 2019.

[114] T.V.M. Santos, M.F.V. Pereira, H.M.N.T. Avelino, F.J.P. Caetano, and J.M.N.A. Fareleira. Viscosity and density measurements on liquid n-tetradecane at moderately high pressures. *Fluid Phase Equilibria*, 453(46):46–57, 2017.

[115] M. Silberberg and P. Amateis. *Chemistry: The Molecular Nature of Matter and Change (8th edition)*. McGraw-Hill Education, New York, 2017.

[116] M.O. Steinhauser and S.Hiermaier. A review of computational methods in materials science: Examples from shock-wave and polymer physics. *International Journal of Molecular Science*, 10:5135–5216, 2009.

[117] H. Sun. COMPASS: An ab initio force-field optimized for condensed-phase applications:overview with details on alkane and benzene compounds. *The Journal of Physical Chemistry B*, 102:7338–7364, 1998.

[118] H. Sun, S.J. Mumby, J.R. Maple, and A.T. Hagler. An ab initio CFF93 all-atom force field for polycarbonates. *Journal of the American Chemical Society*, 16(7):2978–2987, 1994.

[119] T. Suzuki, R.U. Ebert, and G. Schüürmann. Application of neural networks to modeling and estimating temperature-dependent liquid viscosity of organic compounds. *Journal of Chemical Information and Computer Sciences*, 41(3):776–790, 2001.

[120] A. Tarjomannejad. Prediction of the liquid vapor pressure using the artificial neural network-group contribution method. *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, 34(4):97–111, 2015.

[121] M.E. Tuckerman, Y. Liu, G. Ciccotti, and Glenn G.J. Martyna. Non-hamiltonian molecular dynamics: Generalizing hamiltonian phase space principles to non-hamiltonian systems. *The Journal of Chemical Physics*, 115(4):1678–1702, 2001.

[122] O.T. Unke, S. Chmiela, H.E. Sauceda, M. Gastegger, I. Poltavsky, K.T. Schütt, A. Tkatchenko, and K.R. Müller. Machine learning force fields. "https://arxiv.org/abs/2010.07067".

[123] A. Voulodimos, N. Doulamis, A.Doulamis, and E. Protopapadakis. Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2(3): 258–261, 2018.

[124] William W.G. Hoover, D.J. Evans, R.B. Hickman, A.J.C. Ladd, W.T. Ashurst, and B. Moran. Lennard-jones triple-point bulk and shear viscosities. green-kubo theory, hamiltonian mechanics, and nonequilibrium molecular dynamics. *Phys. Rev. A*, 22(4): 1690–1697, 1980.

[125] C. Wohlfarth. *Viscosity of the mixture (1) tridecane; (2) 2,2,4,4,6,8,8-heptamethylnonane*, volume 25. 2009. doi: 10.1007/978-3-540-75486-2_1867.

[126] J. Wu, Z. Shan, and A.F.A. Asfour. Viscometric properties of multicomponent liquid n-alkane systems. *Fluid Phase Equilibria*, 143(1):263–274, 1998.

[127] Y. Yang, T.A. Pakkanen, and R.L. Rowley. Nonequilibrium molecular dynamics simulations of shear viscosity: Isoamyl alcohol, n-butyl acetate, and their mixtures. *International Journal of Thermophysics*, 21(3):703–717, 2000.

[128] Y. Yang, T.A. Pakkanen, and R.L. Rowley. NEMD simulations of viscosity and viscosity index for lubricant-size model molecules. *International Journal of Thermophysics*, 23(6):1441–1454, 2002.

[129] H. Zhang and J.F. Ely. AUA model NEMD and EMD simulations of the shear viscosity of alkane and alcohol systems. *Fluid Phase Equilibria*, 217(1):111–118, 2004.

Appendix A

Chapter 2: Full results

A.1 Heat capacity

Name	$C_{\text{mol}}^{\text{exp}}(\frac{J}{molK})$	$C_{\text{mol}}^{\text{ga}}(\frac{J}{molK})$	$C_{\text{mol}}^{\text{reg}}(\frac{J}{molK})$	$C_{\text{mol}}^{\text{nn}}(\frac{J}{molK})$	$\delta C_{\text{mol}}^{\text{reg}}(\frac{J}{molK})$	$\delta C_{\text{mol}}(\frac{J}{molK})$
2-methylbutane	164.85	163.49	152.93	150.17	4.90	0.68
Pentane	167.19	165.26	153.14	152.38	3.55	0.57
2,2-dimethylbutane	189.67	186.20	186.00	184.14	5.34	0.78
3-methylpentane	191.16	193.28	186.73	183.78	5.28	0.55
2-methylpentane	194.19	193.28	184.69	188.86	5.16	2.14
Hexane	195.52	195.05	184.89	190.66	3.80	1.17
3-methylhexane	216.70	223.07	217.56	218.49	5.69	1.45
2,3-dimethylpentane	219.58	221.31	217.28	218.31	5.56	1.59
2,2-dimethylpentane	221.00	215.99	217.04	217.00	5.79	0.35
2-methylhexane	222.92	223.07	218.22	221.19	4.61	2.32
Heptane	224.64	224.83	218.01	216.02	3.41	0.92
3-ethyl-3-methylpentane	245.89	251.10	247.83	250.19	7.63	1.10
3,3-dimethylhexane	246.60	245.77	248.12	248.67	7.14	1.42
3,4-dimethylhexane	246.90	245.77	248.41	248.94	6.87	1.35
2,3-dimethylhexane	248.78	251.10	248.32	251.46	6.01	2.93
3-ethyl-2-methylpentane	248.91	251.10	248.10	248.23	6.28	0.87
2,5-dimethylhexane	249.20	251.10	248.80	250.33	5.56	1.41
2,2-dimethylhexane	249.20	245.77	247.45	250.37	7.11	0.51
2,4-dimethylhexane	250.08	251.10	248.32	249.09	6.58	0.64
3-methylheptane	250.20	252.86	249.21	249.60	5.25	1.09

3-ethylhexane	250.29	252.86	249.71	247.23	6.03	0.51
4-methylheptane	251.08	252.86	248.77	251.56	6.27	1.13
2-methylheptane	252.00	252.86	250.08	251.24	4.78	1.60
Octane	254.09	254.62	249.29	254.21	3.64	1.39
3,4-dimethylheptane	273.20	280.89	279.77	282.64	6.71	1.76
2,3-dimethylheptane	276.50	280.89	279.35	280.76	6.46	0.33
2,5-dimethylheptane	278.30	280.89	279.84	280.95	6.01	0.48
3,3-diethylpentane	278.80	275.56	278.70	281.96	8.88	0.85
3-ethyl-3-methylhexane	279.60	275.56	278.49	281.03	8.23	0.65
3-ethyl-4-methylhexane	279.90	280.89	279.49	280.60	6.99	0.73
3-methyloctane	281.00	282.65	280.53	281.67	6.22	0.95
4-methyloctane	281.20	282.65	280.20	281.02	5.90	0.59
4-ethylheptane	282.30	282.65	279.62	282.38	7.75	0.78
3,3-dimethylheptane	282.40	275.56	278.43	279.72	7.91	0.87
2,6-dimethylheptane	282.60	280.89	280.24	281.39	6.58	0.53
3-ethyl-2-methylhexane	282.80	280.89	279.13	282.11	6.73	0.50
2,2-dimethylheptane	283.40	275.56	279.31	280.63	7.45	0.80
3,5-dimethylheptane	283.70	280.89	279.49	282.84	7.32	1.72
2-methyloctane	284.20	282.65	281.08	282.95	5.53	0.29
Nonane	284.50	284.41	281.29	288.63	4.22	2.56
4,4-dimethylheptane	285.70	275.56	278.74	280.24	8.07	1.41
2,4-dimethylheptane	287.60	280.89	279.60	283.07	6.23	1.65
4-ethyl-2-methylhexane	289.10	280.89	279.80	282.09	6.56	0.44
3,4-dimethyloctane	303.60	310.68	310.79	311.19	7.18	0.17
4,5-dimethyloctane	303.90	310.68	310.35	312.66	8.39	0.65
3,6-dimethyloctane	305.90	310.68	310.97	312.81	7.58	0.52
2,3-dimethyloctane	307.30	310.68	310.59	312.78	7.70	0.79
2,5-dimethyloctane	309.40	310.68	310.85	314.47	7.37	1.30
3-methylnonane	309.40	312.44	312.01	313.12	6.48	0.80
3-ethyl-3-methylheptane	309.60	305.35	310.10	312.26	8.46	0.99
2,6-dimethyloctane	310.10	310.68	311.58	311.53	6.25	0.93
4-ethyl-3-methylheptane	310.40	310.68	310.50	311.18	7.46	0.94
3-ethyl-4-methylheptane	310.60	310.68	310.77	311.50	7.71	0.50
5-ethyl-2-methylheptane	310.70	310.68	311.05	314.55	6.78	0.33

4-ethyl-4-methylheptane	312.70	305.35	309.88	312.09	8.68	0.85
3,3-diethylhexane	312.90	305.35	310.62	311.83	8.15	0.72
3-ethyl-2-methylheptane	313.10	310.68	310.24	312.71	8.21	0.83
2,7-dimethyloctane	313.30	310.68	311.72	311.73	6.84	0.85
3,3-dimethyloctane	313.30	305.35	310.07	314.57	8.19	1.42
3-ethyloctane	313.30	312.44	312.15	312.29	7.10	0.86
2-methylnonane	313.50	312.44	312.36	313.71	5.78	0.78
4-ethyloctane	313.50	312.44	311.68	312.28	7.84	0.42
2,2-dimethyloctane	314.00	305.35	309.55	312.89	8.17	0.73
4-propylheptane	314.10	312.44	311.29	314.38	8.89	1.26
5-methylnonane	314.40	312.44	310.81	313.32	7.25	0.68
Decane	314.50	314.20	313.15	317.46	4.01	0.62
3,5-dimethyloctane	314.80	310.68	310.53	310.11	7.85	0.71
4,4-dimethyloctane	315.70	305.35	310.05	312.81	7.52	0.84
3,4-diethylhexane	317.30	310.68	310.49	311.99	7.99	0.73
4-methylnonane	317.40	312.44	311.48	312.40	6.12	0.49
2,4-dimethyloctane	318.00	310.68	310.42	311.82	7.63	0.18
4-ethyl-2-methylheptane	320.20	310.68	310.41	313.67	6.95	0.29
6-ethyl-3-methyloctane	338.00	340.46	342.28	340.81	8.42	0.42
3,7-dimethylnonane	339.00	340.46	342.53	341.40	6.90	0.72
3,6-dimethylnonane	339.00	340.46	342.23	341.95	8.19	1.22
3,5-diethylheptane	339.00	340.46	342.15	340.81	8.15	0.57
6-ethyl-2-methyloctane	340.00	340.46	342.59	341.89	7.94	0.25
4-ethyl-2-methyloctane	340.00	340.46	341.73	341.33	8.47	0.08
5-ethyl-3-methyloctane	340.00	340.46	341.84	341.61	8.68	1.06
4-ethyl-3-methyloctane	340.00	340.46	341.52	341.05	7.93	0.35
3,4-dimethylnonane	340.00	340.46	341.35	342.47	8.44	0.49
3-ethyl-3-methyloctane	340.00	340.46	341.41	341.16	9.22	0.27
3-ethyl-4-methyloctane	340.00	340.46	341.57	341.47	9.26	0.52
4,5-dimethylnonane	340.00	340.46	341.27	343.28	8.86	0.87
2,7-dimethylnonane	341.00	340.46	342.83	342.74	6.47	0.90
5-ethyl-2-methyloctane	341.00	340.46	341.69	342.23	7.18	0.62
2,3-dimethylnonane	341.00	340.46	341.57	341.59	8.22	0.70
3-methyl-4-propylheptane	341.00	340.46	341.23	339.72	8.21	1.87

3,5-dimethylnonane	341.00	340.46	341.94	341.13	8.43	1.34
3,3-dimethylnonane	341.00	340.46	341.56	341.88	7.90	0.18
3,4-diethylheptane	341.00	340.46	341.91	341.99	9.63	0.53
2,6-dimethylnonane	342.00	340.46	342.33	342.09	7.63	1.02
2,5-dimethylnonane	342.00	340.46	342.35	342.99	6.97	0.83
3-ethyl-2-methyloctane	342.00	340.46	341.60	342.44	8.50	0.44
3-ethylnonane	342.00	342.23	343.25	343.17	7.80	0.75
3,3-diethylheptane	342.00	335.14	341.62	341.99	9.87	0.44
4,6-dimethylnonane	342.00	340.46	342.11	341.83	8.65	0.64
4-ethyl-4-methyloctane	342.00	335.14	340.86	341.88	9.20	0.21
4-ethylnonane	342.00	342.23	342.91	342.64	7.83	0.88
5-ethylnonane	342.00	342.23	342.97	342.86	7.54	0.36
2-methyl-4-propylheptane	343.00	340.46	341.89	343.71	8.84	0.52
2,2-dimethylnonane	343.00	335.14	341.27	343.52	8.49	0.50
3-methyldecane	343.00	342.23	343.05	343.88	5.92	0.48
4-methyldecane	343.00	342.23	343.23	341.30	6.31	1.35
4,4-dimethylnonane	343.00	335.14	341.49	340.82	8.08	1.68
4-propyloctane	343.00	342.23	343.76	342.29	7.89	0.66
5-methyldecane	343.00	342.23	343.09	341.97	7.84	0.16
5,5-dimethylnonane	343.00	335.14	340.62	340.08	9.14	1.01
2,8-dimethylnonane	344.00	340.46	342.93	339.62	7.24	1.86
2,4-dimethylnonane	344.00	340.46	342.11	342.97	7.22	1.58
4,4-diethylheptane	344.00	335.14	341.51	341.23	10.05	1.10
Undecane	345.00	343.99	343.85	347.78	4.72	0.60
2-methyldecane	345.00	342.23	343.85	342.14	5.44	0.26
4-methyl-4-propylheptane	345.00	335.14	341.64	346.00	9.77	1.34
3,6-diethyloctane	367.00	370.25	373.56	369.90	9.24	1.21
7-ethyl-3-methylnonane	368.00	370.25	374.04	370.92	8.76	0.60
3,8-dimethyldecane	369.00	370.25	373.53	370.49	7.87	0.79
6-ethyl-3-methylnonane	369.00	370.25	372.68	368.66	7.82	2.11
3,7-dimethyldecane	370.00	370.25	373.98	372.16	8.44	0.95
3,6-dimethyldecane	370.00	370.25	373.30	371.60	7.62	0.30
3,4-dimethyldecane	370.00	370.25	372.42	370.85	8.00	0.51
3,5-diethyloctane	370.00	370.25	372.76	371.20	8.68	0.74

4,7-dimethyldecane	370.00	370.25	372.86	370.01	7.74	0.40
7-ethyl-2-methylnonane	371.00	370.25	373.22	372.08	7.17	0.36
6-ethyl-2-methylnonane	371.00	370.25	372.97	371.97	7.40	0.92
5-ethyl-2-methylnonane	371.00	370.25	373.21	371.44	8.73	0.28
3-methyl-5-propyloctane	371.00	370.25	373.52	371.37	9.56	0.39
5-ethyl-3-methylnonane	371.00	370.25	373.46	373.10	9.24	0.87
3-methyl-4-propyloctane	371.00	370.25	372.24	371.26	8.69	0.36
4-ethyl-3-methylnonane	371.00	370.25	372.46	371.37	9.48	0.30
3-ethyl-3-methylnonane	371.00	364.93	371.96	371.55	8.50	0.17
3,3-dimethyldecane	371.00	364.93	372.57	368.76	8.37	2.12
3-ethyl-5-methylnonane	371.00	370.25	373.04	369.61	8.40	2.26
3,4-diethyloctane	371.00	370.25	372.52	367.92	8.93	2.84
3-ethyl-4-methylnonane	371.00	370.25	372.93	370.57	9.51	0.65
6-ethyl-4-methylnonane	371.00	370.25	373.01	371.16	9.43	0.53
5-ethyl-4-methylnonane	371.00	370.25	373.34	370.37	9.71	0.54
4,5-dimethyldecane	371.00	370.25	372.11	371.81	9.17	0.46
4-ethyl-5-methylnonane	371.00	370.25	372.72	372.70	9.72	0.66
5,6-dimethyldecane	371.00	370.25	373.45	372.18	9.61	0.19
2,8-dimethyldecane	372.00	370.25	374.09	371.73	6.69	0.55
2,7-dimethyldecane	372.00	370.25	373.74	372.80	7.65	0.65
2-methyl-5-propyloctane	372.00	370.25	373.64	372.34	9.09	0.73
2,6-dimethyldecane	372.00	370.25	373.61	372.98	7.19	0.59
2,5-dimethyldecane	372.00	370.25	373.15	370.46	8.20	0.41
3-ethyl-2-methylnonane	372.00	370.25	372.58	372.47	9.02	0.31
2,3-dimethyldecane	372.00	370.25	372.47	369.49	7.81	2.24
3,5-dimethyldecane	372.00	370.25	373.40	367.33	8.92	4.14
3-ethyl-4-propylheptane	372.00	370.25	373.14	368.20	10.40	2.85
4-methyl-5-propyloctane	372.00	370.25	371.93	371.17	8.74	0.60
4,6-dimethyldecane	372.00	370.25	373.57	372.12	9.15	0.66
4,5-diethyloctane	372.00	370.25	372.45	371.16	9.12	0.79
3-methylundecane	373.00	372.02	374.97	372.17	7.00	0.97
3-ethyldecane	373.00	372.02	374.83	371.95	6.92	1.43
3,3-diethyloctane	373.00	364.93	372.44	372.51	10.36	0.21
4-ethyl-4-methylnonane	373.00	364.93	372.22	371.92	8.84	1.36

4,4-dimethyldecane	373.00	364.93	371.81	372.31	9.45	0.37
4-ethyldecane	373.00	372.02	374.23	367.75	7.31	2.94
4-propylnonane	373.00	372.02	374.63	368.39	9.25	3.53
5-ethyl-5-methylnonane	373.00	364.93	373.22	371.75	10.17	1.54
5,5-dimethyldecane	373.00	364.93	372.44	372.34	8.74	1.00
5-ethyldecane	373.00	372.02	374.77	371.26	8.98	0.60
5-propylnonane	373.00	372.02	374.68	371.68	9.87	0.52
2,9-dimethyldecane	374.00	370.25	374.34	373.30	7.10	0.81
2-methyl-4-propyloctane	374.00	370.25	373.03	373.84	9.31	1.64
4-ethyl-2-methylnonane	374.00	370.25	373.29	372.27	9.02	0.46
2,4-dimethyldecane	374.00	370.25	372.85	373.82	8.47	0.58
2,2-dimethyldecane	374.00	364.93	372.64	372.50	8.19	0.79
4-methylundecane	374.00	372.02	374.66	372.30	7.48	0.37
5-methylundecane	374.00	372.02	374.18	372.30	6.97	0.65
6-methylundecane	374.00	372.02	373.46	371.62	7.41	0.52
4-methyl-4-propyloctane	375.00	364.93	371.87	371.62	10.00	0.44
4,4-diethyloctane	375.00	364.93	373.26	372.25	10.60	0.28
Dodecane	376.00	373.78	375.62	367.40	5.16	1.84
2-methylundecane	376.00	372.02	375.29	370.99	6.52	1.64
4-ethyl-4-propylheptane	377.00	364.93	372.43	368.37	9.51	2.91

Table A.1 Full molar heat capacity results. The table contains experimental heat capacity values, their group additivity fits [112], and linear regression and neural network cross-validation predictions accompanied by their uncertainties.

A.2 Boiling point

Name	T_b^{exp}(°C)	$T_b^{m7.2}$(°C)	T_b^{reg}(°C)	T_b^{nn}(°C)	δT_b^{reg}(°C)	δT_b^{nn}(°C)
Methane	-161.51		-59.75	-119.77	3.38	4.80
Ethane	-88.58		-34.88	-93.88	3.75	4.28
Propane	-42.07		-24.51	-52.37	6.02	2.93
2-methylpropane	-11.73		3.10	-16.41	9.79	2.81
Butane	-0.50		2.71	-4.70	6.60	2.44

2,2-dimethylpropane	9.50		25.62	13.51	12.31	3.33
2-methylbutane	27.85		27.85	19.31	10.15	1.97
Pentane	36.07		28.90	31.91	7.09	3.50
2,2-dimethylbutane	49.74	50.40	58.95	48.33	8.61	2.64
2,3-dimethylbutane	57.99	57.70	51.16	53.95	12.50	2.57
2-methylpentane	60.27	56.50	56.07	59.30	10.92	1.81
3-methylpentane	63.23	62.60	53.56	57.28	11.80	2.65
Hexane	68.74	66.30	55.60	61.30	7.62	5.87
2,2-dimethylpentane	79.20	76.10	76.39	80.60	14.66	3.80
2,4-dimethylpentane	80.50	80.40	84.86	88.36	8.61	2.63
3,3-dimethylpentane	86.04	85.50	83.37	81.05	9.63	3.16
2,3-dimethylpentane	89.78	88.60	76.80	80.98	13.79	1.15
2-methylhexane	90.05	87.00	82.55	89.67	11.49	3.04
3-methylhexane	91.35	91.30	81.48	86.77	12.50	3.10
3-ethylpentane	93.48	96.40	81.53	85.69	14.31	2.63
Heptane	98.42	96.90	83.03	94.20	8.45	5.33
2,2-dimethylhexane	106.84	103.40	101.78	109.21	15.83	2.17
2,5-dimethylhexane	109.11	107.00	105.35	115.40	13.82	1.09
2,4-dimethylhexane	109.43	110.60	104.12	114.24	14.44	1.72
3,3-dimethylhexane	111.97	110.80	101.37	107.69	16.19	1.63
2,3-dimethylhexane	115.61	113.30	102.67	108.29	14.96	1.07
3-ethyl-2-methylpentane	115.66	117.70	101.11	111.09	15.95	3.14
2-methylheptane	117.65	114.90	112.54	119.08	7.76	2.68
4-methylheptane	117.71	116.90	106.22	116.61	14.00	0.87
3,4-dimethylhexane	117.73		102.78	115.07	15.24	1.61
3-ethyl-3-methylpentane	118.27	119.40	106.47	114.38	10.84	2.01
3-ethylhexane	118.34	121.20	107.12	119.20	14.83	1.97
3-methylheptane	118.93	118.80	112.09	120.82	8.42	1.24
Octane	125.67	126.20	109.01	124.07	8.69	3.21
2,2-dimethylheptane	132.82	129.30	127.47	132.29	16.76	1.96
2,4-dimethylheptane	132.89	133.60	132.52	136.12	10.08	2.06
4-ethyl-2-methylhexane	133.80	136.00	127.60	137.54	16.51	2.29
4,4-dimethylheptane	134.90	133.20	127.14	134.81	17.72	2.22
2,6-dimethylheptane	135.22	133.20	131.27	140.70	14.10	2.12

3,5-dimethylheptane	135.70	138.60	129.52	139.06	16.12	1.20
2,5-dimethylheptane	136.00	135.70	130.46	138.78	15.17	2.44
3,3-dimethylheptane	137.02	135.40	127.41	137.14	17.48	0.45
3-ethyl-4-methylhexane	140.40	120.90	126.74	138.52	17.60	0.93
3,4-dimethylheptane	140.40	120.90	128.29	137.19	16.74	2.26
2,3-dimethylheptane	140.50	138.20	128.26	135.72	16.40	1.86
4-ethylheptane	141.20	143.00	134.85	138.90	16.71	1.16
4-methyloctane	142.44	142.90	133.80	144.60	14.73	2.33
2-methyloctane	143.16	142.50	135.65	142.86	12.33	1.74
3-ethylheptane	143.22	145.20	135.05	143.47	15.65	1.86
3-methyloctane	144.21	145.20	134.41	145.24	13.44	2.62
3,3-diethylpentane	146.19	151.60	126.83	145.54	18.98	2.12
Nonane	150.82	151.60	139.21	147.58	6.31	2.53
2,4-dimethyloctane	155.90	157.80	155.50	159.60	16.68	0.75
4-ethyl-2-methylheptane	156.20		153.08	159.43	17.01	1.11
2,2-dimethyloctane	156.90	155.60	153.07	159.35	18.01	0.95
4,4-dimethyloctane	157.50	156.10	153.04	162.34	19.13	1.69
4-propylheptane	157.50	156.10	161.06	162.06	19.25	1.96
5-ethyl-3-methylheptane	158.20		156.21	161.12	11.58	1.02
2,5-dimethyloctane	158.50	157.90	156.75	159.60	16.15	0.43
3,5-dimethyloctane	159.40	159.80	155.35	160.30	16.87	0.87
5-ethyl-2-methylheptane	159.70		154.09	158.16	17.08	1.69
2,7-dimethyloctane	159.87	159.50	158.05	160.98	15.13	0.98
2,6-dimethyloctane	160.38	160.90	158.43	162.33	10.07	0.63
3,6-dimethyloctane	160.80		156.19	158.41	16.30	0.59
4-ethyl-4-methylheptane	160.80		150.33	161.38	19.71	0.58
3-ethyl-2-methylheptane	161.20	141.80	152.71	158.63	17.97	1.31
3,3-dimethyloctane	161.20	141.80	152.76	157.80	18.52	1.43
4,5-dimethyloctane	162.13	161.40	154.41	159.68	17.97	0.21
4-ethyl-3-methylheptane	162.20		151.86	158.47	18.58	0.31
3-ethyl-4-methylheptane	163.00	138.70	154.99	162.77	19.10	1.69
3,4-dimethyloctane	163.40		153.89	162.03	18.04	1.13
4-ethyloctane	163.64	165.50	160.04	161.98	10.97	0.78
3-ethyl-3-methylheptane	163.80		151.29	159.03	18.67	1.60

3,4-diethylhexane	163.90		154.86	161.50	12.73	1.03
2,3-dimethyloctane	164.31	162.70	155.32	159.95	11.18	0.52
5-methylnonane	165.10	165.10	159.20	164.54	16.13	1.61
4-methylnonane	165.70	165.20	160.30	165.54	15.35	0.60
3,3-diethylhexane	166.30		152.20	161.18	20.39	1.66
3-ethyloctane	166.50	169.10	161.02	163.97	15.92	0.89
2-methylnonane	167.00	166.20	162.01	166.04	13.18	1.41
3-methylnonane	167.30	168.70	161.09	167.93	13.93	1.02
2-methyl-4-propylheptane	174.00		176.61	178.34	19.59	1.12
Decane	174.16	174.40	162.42	168.85	9.75	0.49
4-ethyl-2-methyloctane	176.00		179.66	181.67	18.52	0.79
3-ethyl-5-methyloctane	176.00		181.95	180.20	19.64	0.82
5,5-dimethylnonane	177.00		178.67	178.35	20.78	2.17
2,4-dimethylnonane	177.40		181.32	182.54	17.44	0.79
5-ethyl-2-methyloctane	178.00		180.73	180.44	17.89	0.40
4-methyl-4-propylheptane	178.00		173.90	177.40	21.55	2.93
4,4-dimethylnonane	178.00		179.43	179.85	19.64	0.68
2,6-dimethylnonane	179.00		183.68	180.77	16.70	1.03
2,5-dimethylnonane	179.00		182.45	180.41	17.32	0.68
3-methyl-4-propylheptane	179.00		175.62	179.76	19.88	0.80
3,5-diethylheptane	179.00		180.23	184.19	20.31	0.65
4,6-dimethylnonane	179.00		181.36	182.89	12.13	1.75
2,2-dimethylnonane	180.00		179.22	182.99	19.11	0.81
3,5-dimethylnonane	180.00		180.92	184.43	18.45	1.53
3,4-diethylheptane	180.00		179.40	183.85	20.62	1.86
4-ethyl-4-methyloctane	180.00		176.66	180.07	21.03	0.75
4-propyloctane	180.00		189.76	180.33	19.80	1.16
6-ethyl-2-methyloctane	182.00		181.81	181.20	17.26	0.91
3-ethyl-2-methyloctane	182.00		178.41	182.27	12.39	0.88
6-ethyl-3-methyloctane	182.00		179.68	179.53	18.07	0.60
3,6-dimethylnonane	182.00		182.15	178.80	17.83	0.76
3,3-dimethylnonane	182.00		179.06	180.29	19.59	0.95
4,5-dimethylnonane	182.00		180.27	182.11	19.07	0.69
2,8-dimethylnonane	183.00		185.68	182.81	15.34	0.63

2,7-dimethylnonane	183.00		183.84	179.83	15.74	1.81
4-ethyl-3-methyloctane	183.00		179.00	186.14	12.68	1.78
5-ethylnonane	183.00		185.49	186.73	18.68	1.42
4,4-diethylheptane	183.30		178.14	181.62	22.37	1.18
3,7-dimethylnonane	184.00		182.58	182.90	17.39	0.81
4-ethylnonane	184.00		187.17	184.13	17.86	0.98
3,4-dimethylnonane	185.00		179.74	182.49	12.21	0.86
3-ethyl-3-methyloctane	185.70		176.42	179.57	20.56	1.72
2,3-dimethylnonane	186.00		180.00	187.26	18.55	0.86
5-methyldecane	186.00		185.90	182.84	16.66	1.43
3,3-diethylheptane	186.90		177.90	182.76	21.90	2.26
4-methyldecane	187.90		186.78	187.78	15.91	0.87
3-ethylnonane	188.00		185.36	182.13	10.68	1.50
2-methyldecane	189.19		190.71	188.40	13.56	0.74
3-methyldecane	190.30		189.18	188.08	14.69	1.09
2-methyl-5-propyloctane	191.00		203.58	196.16	12.86	0.92
2-methyl-4-propyloctane	191.00		202.54	194.41	13.23	0.39
3-methyl-5-propyloctane	194.00		203.13	196.15	13.52	0.68
5-ethyl-3-methylnonane	195.00		204.70	196.04	19.74	1.30
4-methyl-4-propyloctane	195.00		201.65	199.08	14.55	0.65
Undecane	195.93		188.94	191.60	7.04	1.13
5-ethyl-2-methylnonane	196.00		205.65	199.13	18.64	0.52
4-ethyl-2-methylnonane	196.00		205.53	198.46	19.69	0.62
3-ethyl-4-propylheptane	196.00		202.62	194.76	23.19	0.86
6-ethyl-4-methylnonane	196.00		204.25	198.69	20.99	1.28
4-methyl-5-propyloctane	196.00		201.38	199.12	22.33	0.71
3-methyl-4-propyloctane	197.00		202.10	196.34	13.89	0.58
3,5-diethyloctane	197.00		206.10	199.75	21.48	0.83
4,5-diethyloctane	197.00		205.15	197.96	22.25	0.82
4-ethyl-4-propylheptane	197.00		202.07	200.79	23.88	1.38
5,5-dimethyldecane	197.00		204.60	195.25	21.73	1.29
5-propylnonane	197.00		215.15	197.81	21.48	2.34
6-ethyl-2-methylnonane	198.00		205.36	202.74	12.02	0.80
2,6-dimethyldecane	198.00		209.40	202.22	17.77	0.82

2,5-dimethyldecane	198.00		208.15	199.99	18.49	0.90
4,6-dimethyldecane	198.00		206.98	197.26	19.60	0.88
5-ethyl-4-methylnonane	198.00		203.42	201.22	13.71	0.69
5-ethyl-5-methylnonane	198.00		202.81	198.53	21.94	0.93
6-ethyl-3-methylnonane	199.00		204.91	199.99	12.68	1.17
3-ethyl-5-methylnonane	199.00		207.82	200.24	20.82	0.62
3,4-diethyloctane	199.00		205.54	200.05	22.28	1.01
4,7-dimethyldecane	199.00		206.23	202.60	12.49	1.59
4-ethyl-4-methylnonane	199.00		203.05	200.19	21.91	1.54
4,4-dimethyldecane	199.00		204.77	197.81	21.25	0.96
4-ethyl-5-methylnonane	199.00		207.37	201.04	22.12	0.75
4-propylnonane	199.00		214.67	201.65	20.05	0.75
2,4-dimethyldecane	200.00		205.52	201.44	19.31	0.31
3,5-dimethyldecane	200.00		207.08	201.69	19.07	0.37
3-ethyl-6-methylnonane	200.00		208.90	199.04	20.19	0.97
4,4-diethyloctane	200.00		202.08	198.61	14.86	1.58
2,2-dimethyldecane	201.00		203.32	199.21	20.53	0.96
3,6-dimethyldecane	201.00		207.84	200.96	19.00	0.80
4-ethyl-3-methylnonane	201.00		204.29	198.50	20.80	0.54
5,6-dimethyldecane	201.00		204.78	200.91	21.26	1.44
7-ethyl-2-methylnonane	202.00		206.39	200.27	11.65	1.07
2,7-dimethyldecane	202.00		209.71	203.23	16.84	0.73
3,7-dimethyldecane	202.00		208.76	203.08	17.94	0.80
3,6-diethyloctane	202.00		208.00	198.64	21.04	1.58
4,5-dimethyldecane	202.00		205.09	202.62	20.75	0.78
5-ethyldecane	202.00		209.33	201.77	12.36	0.34
3-ethyl-2-methylnonane	203.00		202.24	201.77	13.13	0.81
7-ethyl-3-methylnonane	203.00		206.38	201.69	18.61	0.95
3,3-dimethyldecane	203.00		204.16	197.86	20.85	0.78
3-ethyl-4-methylnonane	203.00		207.64	198.15	21.05	0.76
2,9-dimethyldecane	204.00		213.06	203.40	16.02	0.54
2,8-dimethyldecane	204.00		211.83	199.50	16.64	0.92
3-ethyl-3-methylnonane	204.00		203.21	203.11	21.43	0.77
4-ethyldecane	204.00		209.77	204.27	11.70	1.03

3,8-dimethyldecane	205.00		210.31	203.21	17.62	0.70
3,4-dimethyldecane	205.00		203.57	201.83	12.94	0.20
3,3-diethyloctane	205.00		204.31	198.66	22.89	1.37
2,3-dimethyldecane	206.00		206.17	201.52	19.66	1.03
5-methylundecane	206.00		212.61	205.59	17.20	0.98
6-methylundecane	206.00		211.51	204.81	18.64	0.70
3-ethyldecane	208.70		214.51	198.64	17.34	2.08
4-methylundecane	209.00		211.12	206.17	10.55	1.44
2-methylundecane	210.00		216.50	209.29	13.97	0.96
3-methylundecane	210.30		216.11	208.84	15.24	0.67
Dodecane	216.32		215.46	206.20	11.27	2.36

Table A.2 Full boiling point results. The table contains boiling point values, their model 7.2 fits [16], and linear regression and neural network cross-validation predictions accompanied by their uncertainties.

A.3 Antoine coefficients

Name	$B^{\text{exp}}(kPa^{\circ}C)$	$B^{\text{reg}}(kPa^{\circ}C)$	$B^{\text{nn}}(kPa^{\circ}C)$	$\delta B^{\text{reg}}(kPa^{\circ}C)$	$\delta B^{\text{nn}}(kPa^{\circ}C)$
Methane	500.25	666.12	546.43	22.69	8.31
Ethane	657.10	747.70	670.64	36.09	3.19
Propane	828.30	840.08	801.28	43.43	2.95
Butane	960.05	921.85	935.32	44.98	6.91
Pentane	1082.54	1027.57	1060.19	44.58	7.80
2,2-dimethylbutane	1090.16	1110.98	1093.30	95.78	11.76
2,3-dimethylbutane	1127.40	1111.02	1119.68	87.84	8.76
Hexane	1143.56	1117.93	1165.59	50.05	11.35
2-methylpentane	1145.80	1119.23	1142.31	72.87	10.02
3-methylpentane	1162.37	1134.73	1131.10	82.06	7.88
2,2-dimethylpentane	1191.96	1186.60	1197.65	105.25	10.22
2,4-dimethylpentane	1193.61	1206.92	1223.90	95.11	12.78
3,3-dimethylpentane	1227.02	1195.10	1236.69	132.55	7.36
2-methylhexane	1235.52	1224.42	1243.40	86.70	5.72

2,3-dimethylpentane	1238.99	1203.31	1210.87	80.21	13.34
3-methylhexane	1242.02	1214.43	1251.60	112.87	11.92
3-ethylpentane	1254.06	1216.92	1249.53	124.69	10.58
2,2-dimethylhexane	1271.18	1274.57	1290.53	156.19	7.51
Heptane	1271.23	1208.30	1253.91	55.53	8.55
2,5-dimethylhexane	1285.47	1303.35	1310.16	141.23	6.28
2,5-dimethylhexane	1285.85	1302.23	1307.17	100.28	5.12
3,3-dimethylhexane	1306.96	1285.56	1316.72	116.41	4.27
2,3-dimethylhexane	1314.29	1280.55	1298.51	111.41	8.76
3-ethyl-2-methylpentane	1317.05	1293.61	1316.47	155.92	9.35
4-methylheptane	1325.74	1318.94	1318.53	96.86	2.90
3-methylheptane	1326.14	1311.48	1325.46	124.93	3.24
3-ethylhexane	1327.93	1312.03	1343.39	101.22	16.20
3,4-dimethylhexane	1329.40	1295.96	1310.16	87.97	5.24
2-methylheptane	1335.22	1313.08	1332.71	126.54	9.99
3-ethyl-3-methylpentane	1345.92	1296.92	1324.82	95.08	6.81
2,2-dimethylheptane	1346.10	1362.03	1374.77	173.23	4.36
Octane	1347.92	1302.02	1355.19	80.14	6.64
4,4-dimethylheptane	1360.20	1384.51	1389.40	127.57	4.49
3,3-dimethylheptane	1365.40	1368.63	1374.47	100.41	2.15
2,4-dimethylheptane	1367.00	1375.19	1393.45	94.64	4.69
2,6-dimethylheptane	1376.40	1392.88	1379.09	87.68	3.80
4-ethyl-2-methylhexane	1377.80	1385.00	1398.31	98.26	7.28
3,5-dimethylheptane	1378.60	1390.84	1391.98	120.51	3.93
2,5-dimethylheptane	1380.00	1384.66	1381.36	112.49	3.53
3-ethyl-3-methylhexane	1389.70	1379.46	1401.14	171.36	3.82
4-methyloctane	1395.90	1414.60	1388.00	106.79	2.83
3-ethyl-2-methylhexane	1397.40	1378.82	1376.11	130.11	6.81
2-methyloctane	1399.90	1409.19	1405.75	101.38	11.24
3,4-dimethylheptane	1400.50	1377.48	1383.49	96.93	4.68
2,3-dimethylheptane	1401.00	1364.67	1369.12	126.02	3.16
3-methyloctane	1404.30	1410.19	1402.66	95.68	3.85
4-ethylheptane	1406.20	1414.80	1403.82	90.00	4.65
3-ethyl-4-methylhexane	1408.50	1393.77	1397.58	125.61	5.36

3-ethylheptane	1415.70	1411.02	1398.51	148.81	5.42
Nonane	1432.28	1391.61	1424.58	75.75	7.79
4-ethyl-2-methylheptane	1438.20	1471.20	1457.11	137.46	4.96
2,7-dimethyloctane	1444.19	1492.26	1448.16	121.04	6.32
4,4-dimethyloctane	1446.90	1468.31	1464.46	138.80	5.07
3,5-dimethyloctane	1448.80	1485.23	1462.34	126.39	3.17
5-ethyl-2-methylheptane	1448.80	1477.04	1450.19	127.69	6.69
2,6-dimethyloctane	1450.53	1487.86	1452.25	170.33	6.07
5-ethyl-3-methylheptane	1455.20	1489.38	1458.58	131.83	4.40
3,3-diethylpentane	1455.75	1391.22	1422.25	137.37	11.25
3,6-dimethyloctane	1456.90	1486.25	1454.81	168.72	4.17
3,3-dimethyloctane	1457.60	1453.14	1454.76	138.87	6.99
3-ethyl-2-methylheptane	1457.90	1464.69	1447.14	137.21	3.65
4-propylheptane	1458.70	1531.14	1463.94	136.53	6.63
4,5-dimethyloctane	1460.52	1474.20	1468.61	134.66	3.05
4-ethyl-3-methylheptane	1464.20	1468.80	1455.68	109.50	3.18
4-ethyl-4-methylheptane	1464.90	1483.25	1474.61	140.74	4.32
3,4-dimethyloctane	1465.90	1466.77	1462.46	131.15	7.65
3-ethyl-4-methylheptane	1466.80	1479.73	1462.04	140.97	3.59
3-ethyl-3-methylheptane	1472.90	1465.40	1466.51	143.13	5.36
3,4-diethylhexane	1473.20	1488.60	1488.83	190.25	4.12
Decane	1479.02	1479.39	1493.46	71.94	7.77
3,3-diethylhexane	1490.90	1481.83	1481.89	150.31	5.49
Undecane	1570.86	1595.66	1559.72	88.08	6.51
Dodecane	1628.18	1700.40	1628.08	164.55	8.49

Table A.3 Full Antoine B coefficient results. The table contains experimental coefficient values and their linear regression and neural network cross-validation predictions accompanied by uncertainties.

Name	$C^{\mathrm{exp}}(^{\circ}C)$	$C^{\mathrm{reg}}(^{\circ}C)$	$C^{\mathrm{nn}}(^{\circ}C)$	$\delta C^{\mathrm{reg}}(^{\circ}C)$	$\delta C^{\mathrm{nn}}(^{\circ}C)$
Dodecane	180.64	176.47	181.87	2.49	0.97
Undecane	187.91	185.35	186.18	2.29	0.39
Decane	191.99	194.42	194.00	1.76	0.31

2,7-dimethyloctane	198.27	198.77	200.16	4.33	0.41
2,6-dimethyloctane	199.40	199.09	200.61	4.00	0.31
3,6-dimethyloctane	200.50	199.86	200.95	4.34	0.28
4-propylheptane	201.20	199.61	201.98	5.67	0.51
5-ethyl-2-methylheptane	201.30	201.44	201.57	4.39	0.28
3,5-dimethyloctane	201.30	200.37	200.84	4.48	0.43
3,4-dimethyloctane	201.40	200.93	202.23	4.70	0.16
4,5-dimethyloctane	201.90	201.24	202.31	5.57	0.55
Nonane	202.04	201.65	201.72	2.10	1.08
3-ethyl-2-methylheptane	202.20	202.56	203.06	4.82	0.66
4-ethyl-2-methylheptane	202.30	202.34	202.10	5.22	0.32
5-ethyl-3-methylheptane	202.40	202.15	202.52	4.73	0.30
3,3-dimethyloctane	202.50	200.99	203.17	5.22	0.68
3-ethyl-4-methylheptane	203.00	201.55	203.55	5.32	0.37
4-ethyl-3-methylheptane	203.30	203.09	203.14	3.72	0.23
4,4-dimethyloctane	203.50	201.37	202.59	5.37	0.16
2-methyloctane	204.00	203.70	203.42	3.51	0.29
4-methyloctane	204.00	205.17	203.74	4.30	0.40
3-methyloctane	204.00	204.37	204.61	3.25	0.44
3,4-diethylhexane	204.20	204.26	205.46	5.32	0.27
3-ethyl-3-methylheptane	204.50	203.20	202.89	5.42	0.48
4-ethyl-4-methylheptane	205.20	203.43	204.38	5.29	0.20
4-ethylheptane	206.00	206.88	206.01	3.14	0.67
3-ethylheptane	206.00	205.91	206.44	3.85	0.12
3,3-diethylhexane	206.20	204.14	206.30	5.75	0.50
2,3-dimethylheptane	208.00	208.28	209.07	4.85	0.79
3,5-dimethylheptane	208.00	208.01	208.71	4.80	0.77
2,6-dimethylheptane	208.00	206.89	207.73	2.75	0.22
2,4-dimethylheptane	208.00	207.93	205.97	3.06	0.93
3,4-dimethylheptane	208.00	208.56	206.54	3.30	0.42
3,3-dimethylheptane	208.00	209.07	208.71	3.46	0.60
2,2-dimethylheptane	208.00	208.99	208.33	4.55	0.05
2,5-dimethylheptane	208.00	207.08	210.00	3.98	0.38
4,4-dimethylheptane	208.00	208.99	207.11	5.01	0.20

Octane	208.70	209.75	209.23	2.35	0.52
3-ethyl-2-methylhexane	209.00	210.34	208.05	5.07	0.38
4-ethyl-2-methylhexane	209.00	209.61	209.14	3.22	0.19
3-ethyl-3-methylhexane	209.00	211.24	210.86	4.88	0.22
3-ethyl-4-methylhexane	209.00	210.19	211.01	4.60	0.28
3-methylheptane	211.81	212.29	212.12	3.18	0.17
4-methylheptane	212.37	213.26	212.53	3.73	0.39
3-ethylhexane	212.65	213.93	213.30	3.65	0.31
2-methylheptane	213.42	211.52	211.63	2.84	0.66
2,3-dimethylhexane	214.06	215.46	214.80	4.09	0.67
2,5-dimethylhexane	214.25	214.76	214.33	3.51	0.29
2,5-dimethylhexane	214.60	214.60	213.53	3.46	0.42
2,2-dimethylhexane	214.83	216.53	215.20	4.19	0.47
3,4-dimethylhexane	214.84	215.71	216.19	3.05	0.44
3-ethyl-2-methylpentane	215.23	218.01	216.42	4.18	0.18
3,3-diethylpentane	216.13	211.71	210.36	5.13	0.58
Heptane	217.23	217.44	215.07	1.46	0.94
3,3-dimethylhexane	217.38	216.22	216.53	4.49	0.84
3-methylhexane	219.44	220.42	219.33	3.05	0.41
2-methylhexane	219.50	219.70	219.70	3.21	0.26
3-ethyl-3-methylpentane	219.58	217.92	217.92	3.36	0.85
3-ethylpentane	220.14	222.17	221.24	3.59	1.43
Hexane	221.16	225.11	225.62	1.36	0.33
2,4-dimethylpentane	221.81	222.69	222.21	3.49	0.35
2,3-dimethylpentane	221.94	222.76	222.59	2.70	0.34
2,2-dimethylpentane	223.50	223.45	223.15	3.96	0.45
3,3-dimethylpentane	225.12	224.26	224.06	3.94	0.81
2-methylpentane	227.82	227.76	224.70	2.52	1.79
3-methylpentane	228.29	228.44	226.13	3.46	1.73
2,3-dimethylbutane	228.97	230.67	231.47	3.20	0.32
2,2-dimethylbutane	230.52	231.27	230.08	3.95	0.58
Pentane	233.90	232.79	231.66	1.26	0.51
Butane	241.54	241.63	242.66	1.46	0.35
Propane	250.18	249.75	251.82	1.61	0.30

Ethane	256.11	257.75	261.14	1.46	0.42
Methane	272.73	263.48	265.12	0.86	0.97

Table A.4 Full Antoine C coefficient results. The table contains experimental coefficient values and their linear regression and neural network cross-validation predictions accompanied by uncertainties.

A.4 Flash Point

Name	$T_f^{exp}(^\circ C)$	$T_f^{gc}(^\circ C)$	$T_f^{lin}(^\circ C)$	$T_f^{nn}(^\circ C)$	$\delta T_f^{lin}(^\circ C)$	$\delta T_f^{nn}(^\circ C)$
Methane	-188.00		-189.86	-183.92	3.21	2.10
Ethane	-135.00	-129.04	-136.22	-133.70	1.96	0.67
Propane	-104.00	-97.15	-102.37	-98.49	4.53	0.08
Butane	-60.00	-71.15	-74.04	-74.27	5.46	3.28
Pentane	-49.00	-47.15	-46.66	-46.13	8.16	0.39
Hexane	-26.00	-26.15	-22.51	-21.69	10.92	0.68
Heptane	-4.00	-6.15	-1.86	-4.41	12.69	0.19
Octane	13.00	11.85	17.77	10.22	14.24	1.37
Nonane	31.00	28.85	35.50	29.36	15.56	2.51
Decane	46.00	44.85	52.59	43.33	17.46	3.09
Undecane	60.00	60.85	67.33	60.71	21.45	0.53
Dodecane	71.00	74.85	82.22	72.80	20.82	0.91
Tridecane	94.00	89.85	94.98	90.28	22.10	0.80
Tetradecane	100.00	102.85	107.88	111.21	23.86	2.37
Pentadecane	132.00	115.85	119.09	123.50	22.67	1.44
Hexadecane	135.00	128.85	131.32	138.17	26.21	0.49
Heptadecane	149.00	141.18	141.52	147.09	33.95	3.07
Octadecane	165.00	153.14	151.50	160.34	22.92	1.12
Nonadecane	168.00	164.79	161.28	163.96	29.07	3.54
Eicosane	176.00	176.13	170.86	175.66	31.79	0.43
Octacosane	227.00	258.19	234.81	225.55	40.83	1.25
Triacontane	238.00	276.80	250.04	239.83	47.75	6.16

Table A.5 Full flash point results. The table contains boiling point values, their group contribution [36] fits, and linear regression and neural network cross-validation predictions accompanied by their uncertainties.

A.5 Melting point

Name	$T_{\mathrm{m}}^{\mathrm{exp}}(^{\circ}C)$	$T_{\mathrm{m}}^{\mathrm{reg}}(^{\circ}C)$	$T_{\mathrm{m}}^{\mathrm{reg}+}(^{\circ}C)$	$\delta T_{\mathrm{m}}^{\mathrm{reg}}(^{\circ}C)$	$\delta T_{\mathrm{m}}^{\mathrm{reg}+}(^{\circ}C)$
Propane	-187.68	-146.71	-165.92	20.68	18.35
Ethane	-182.80	-166.16	-121.87	24.60	19.46
Methane	-182.47	-179.18	-83.40	25.68	17.67
2-methylbutane	-159.90	-133.67	-143.54	44.11	40.07
2-methylpropane	-159.60	-161.84	-168.03	33.22	36.58
2-methylpentane	-153.66	-122.70	-138.59	38.46	42.16
Butane	-138.36	-137.80	-145.47	27.69	20.03
3,3-dimethylpentane	-134.44	-127.87	-133.46	60.14	67.09
Pentane	-129.73	-122.58	-134.24	23.59	22.05
2,3-dimethylbutane	-128.54	-112.40	-117.15	59.18	56.32
3,3-dimethylhexane	-126.10	-126.26	-123.49	69.21	73.30
2,2-dimethylpentane	-123.81	-131.56	-140.57	47.06	47.38
2,2-dimethylhexane	-121.18	-123.07	-123.36	60.41	67.42
4-methylheptane	-120.96	-122.96	-127.13	50.36	56.96
3-methylheptane	-120.53	-107.52	-118.88	43.96	49.16
2,3-dimethylpentane	-119.24	-120.22	-137.49	55.35	60.19
2,4-dimethylpentane	-119.22	-100.17	-120.13	50.36	56.64
3-ethylpentane	-118.58	-138.77	-137.37	50.63	59.41
2-methylhexane	-118.27	-110.36	-118.67	44.42	43.76
2,3-dimethylheptane	-116.70	-113.84	-121.04	51.61	52.86
5-ethyl-2-methylnonane	-116.00	-31.98	-39.02	66.89	83.32
3-ethyl-2-methylpentane	-115.00	-82.79	-96.39	52.61	53.26
4-methyloctane	-113.20	-112.09	-110.01	56.85	60.50
2,2-dimethylheptane	-113.00	-129.46	-132.78	70.59	79.23
4,4-dimethyloctane	-110.00	-128.22	-112.99	81.23	86.48

2-methylheptane	-109.04	-95.26	-105.89	40.25	43.71
3-methyloctane	-107.50	-95.56	-96.28	52.45	52.25
2,6-dimethylheptane	-102.90	-72.34	-86.39	52.61	60.35
2,2-dimethylbutane	-99.87	-129.58	-143.49	57.08	59.84
4-methylnonane	-98.70	-95.78	-97.30	46.03	56.66
Hexane	-95.32	-114.08	-102.36	24.33	26.58
4-methyldecane	-92.00	-78.87	-70.14	58.49	60.70
2,5-dimethylhexane	-91.17	-88.16	-92.55	57.96	59.13
3-ethyl-3-methylpentane	-90.84	-106.42	-123.62	70.49	78.61
Heptane	-90.58	-93.12	-100.37	26.30	24.78
5-methylnonane	-87.70	-112.95	-105.43	57.56	68.87
3-methylnonane	-84.33	-82.68	-87.62	47.94	50.52
2,5-dimethyldecane	-84.00	-85.52	-67.94	76.74	82.90
2-methyloctane	-80.40	-81.82	-82.73	49.34	47.24
3-methyldecane	-79.50	-67.32	-62.84	54.58	53.98
2-methylnonane	-74.65	-67.97	-72.40	42.02	48.60
3-methylundecane	-58.00	-51.05	-53.05	39.86	40.13
Octane	-56.76	-88.04	-72.89	26.57	30.14
2,7-dimethyloctane	-54.00	-69.40	-57.67	64.88	60.65
Nonane	-53.49	-77.77	-76.30	30.77	29.43
2-methyldecane	-48.86	-54.57	-49.48	52.28	50.63
2-methylundecane	-46.81	-40.95	-38.84	47.28	52.32
3,3-diethylpentane	-33.09	-100.71	-17.80	100.09	91.95
Decane	-29.64	-62.00	-43.41	28.81	33.71
Undecane	-25.58	-53.16	-45.30	34.11	33.08
2,2-dimethylpropane	-16.55	-136.70	-105.68	52.78	38.79
Dodecane	-9.58	-40.86	-18.88	42.56	34.91

Table A.6 Full melting point linear regression results.. The table contains experimental melting point values and linear regression cross-validation predictions with uncertainties, both without and with symmetry parameters included.

Name	$T_m^{exp}(^\circ C)$	$T_m^{nn}(^\circ C)$	$T_m^{nn+}(^\circ C)$	$\delta T_m^{nn}(^\circ C)$	$\delta T_m^{nn+}(^\circ C)$
Propane	-187.68	-161.12	-167.48	2.44	3.08

Ethane	-182.80	-179.14	-188.90	3.42	4.95
Methane	-182.47	-165.29	-140.69	15.00	4.22
2-methylbutane	-159.90	-153.94	-156.54	3.93	4.31
2-methylpropane	-159.60	-143.10	-173.38	9.32	2.63
2-methylpentane	-153.66	-143.23	-145.42	5.33	2.84
Butane	-138.36	-150.39	-162.08	1.82	2.98
3,3-dimethylpentane	-134.44	-90.63	-106.48	15.08	5.25
Pentane	-129.73	-135.03	-141.96	4.54	2.45
2,3-dimethylbutane	-128.54	-93.63	-97.84	5.05	2.33
3,3-dimethylhexane	-126.10	-119.61	-106.40	4.62	2.00
2,2-dimethylpentane	-123.81	-125.94	-130.84	2.97	1.86
2,2-dimethylhexane	-121.18	-119.14	-108.39	5.39	4.33
4-methylheptane	-120.96	-115.74	-125.27	4.91	1.67
3-methylheptane	-120.53	-125.84	-124.10	2.77	2.69
2,3-dimethylpentane	-119.24	-112.22	-121.49	6.18	3.17
2,4-dimethylpentane	-119.22	-102.69	-119.97	2.99	2.15
3-ethylpentane	-118.58	-95.68	-128.90	19.99	7.38
2-methylhexane	-118.27	-130.02	-115.66	1.11	4.73
2,3-dimethylheptane	-116.70	-116.39	-118.01	4.66	1.97
5-ethyl-2-methylnonane	-116.00	-101.51	-96.70	12.09	3.09
3-ethyl-2-methylpentane	-115.00	-98.20	-105.13	8.99	4.44
4-methyloctane	-113.20	-114.61	-111.15	2.85	2.20
2,2-dimethylheptane	-113.00	-121.45	-118.57	3.78	2.84
4,4-dimethyloctane	-110.00	-121.56	-102.99	5.90	2.48
2-methylheptane	-109.04	-119.19	-106.11	8.22	3.21
3-methyloctane	-107.50	-107.75	-99.67	3.04	1.34
2,6-dimethylheptane	-102.90	-80.96	-97.35	5.94	2.40
2,2-dimethylbutane	-99.87	-108.65	-135.80	2.57	3.51
4-methylnonane	-98.70	-91.13	-99.36	3.78	3.38
Hexane	-95.32	-101.92	-110.15	2.97	2.99
4-methyldecane	-92.00	-77.50	-88.23	6.16	3.93
2,5-dimethylhexane	-91.17	-101.17	-80.22	1.27	3.77
3-ethyl-3-methylpentane	-90.84	-78.61	-123.25	31.86	4.50
Heptane	-90.58	-80.60	-94.74	6.22	1.05

5-methylnonane	-87.70	-109.27	-103.20	5.45	3.52
3-methylnonane	-84.33	-90.63	-91.41	4.26	3.94
2,5-dimethyldecane	-84.00	-100.30	-103.87	6.96	4.71
2-methyloctane	-80.40	-90.40	-80.36	4.61	0.65
3-methyldecane	-79.50	-67.67	-79.18	1.86	5.26
2-methylnonane	-74.65	-74.83	-68.51	5.67	1.98
3-methylundecane	-58.00	-68.16	-67.70	9.29	1.74
Octane	-56.76	-83.04	-60.06	10.67	2.18
2,7-dimethyloctane	-54.00	-90.50	-79.12	2.19	2.38
Nonane	-53.49	-40.06	-50.75	2.12	4.02
2-methyldecane	-48.86	-59.64	-59.57	3.01	2.77
2-methylundecane	-46.81	-45.94	-46.15	3.32	3.15
3,3-diethylpentane	-33.09	-8.18	26.35	11.93	15.43
Decane	-29.64	-34.52	-26.24	3.35	3.72
Undecane	-25.58	-36.01	-26.26	13.24	4.25
2,2-dimethylpropane	-16.55	-126.15	-90.54	6.31	7.26
Dodecane	-9.58	-1.93	-2.96	4.56	3.92

Table A.7 Full melting point neural network results. The table contains experimental melting point values and neural cross-validation predictions with uncertainties, both without and with symmetry parameters included.

A.6 Kinematic viscosity

Name	$T(^\circ C)$	$\rho^{\exp}(\frac{g}{l})$	$\rho^{\mathrm{reg}}(\frac{g}{l})$	$\rho^{\mathrm{nn}}(\frac{g}{l})$	$\delta\rho^{\mathrm{reg}}(\frac{g}{l})$	$\delta\rho^{\mathrm{nn}}(\frac{g}{l})$
Decane	25.00	726.60	724.76	727.29	4.69	0.64
Decane	50.00	707.40	705.72	705.41	5.00	1.14
Decane	75.00	687.70	686.70	683.60	5.43	1.32
Decane	75.00	687.90	686.05	682.53	5.20	2.98
Decane	100.00	667.50	666.89	663.58	6.05	1.42
Dodecane	20.00	748.50	745.30	748.89	5.04	0.85
Dodecane	25.00	746.00	741.42	745.42	5.44	0.28
Dodecane	30.00	741.40	737.88	741.75	5.08	0.79

Dodecane	40.00	734.10	730.18	734.33	5.42	0.51
Dodecane	50.00	726.90	722.44	727.59	5.41	0.23
Dodecane	50.00	727.40	722.23	727.59	5.29	0.23
Dodecane	60.00	719.70	714.90	720.23	5.42	0.69
Dodecane	70.00	711.90	707.06	713.46	6.06	0.41
Dodecane	75.00	708.90	703.40	708.79	5.72	0.84
Dodecane	80.00	704.40	699.07	705.10	5.66	1.00
Dodecane	90.00	697.00	692.07	697.50	6.03	1.27
Dodecane	100.00	689.80	684.44	687.45	6.15	3.26
Dodecane	100.00	690.00	683.63	689.58	5.90	1.43
Dodecane	125.00	670.90	665.07	667.25	6.82	1.90
Dodecane	150.00	651.10	646.25	647.27	6.65	3.13
Dodecane	175.00	630.40	626.90	626.49	7.51	2.92
Dodecane	200.00	608.90	606.44	605.91	7.13	2.76
Nonane	-17.78	749.99	748.17	746.24	4.21	1.38
Nonane	-12.22	745.34	744.93	743.12	3.93	1.01
Nonane	-6.67	740.54	740.53	739.98	4.09	0.47
Nonane	-1.11	735.89	736.35	735.46	4.07	0.48
Nonane	4.44	731.08	732.32	729.44	4.18	1.07
Nonane	10.00	726.44	726.96	726.64	4.59	0.20
Nonane	15.56	721.63	723.49	720.01	4.28	1.10
Nonane	21.11	716.99	719.37	716.67	4.56	0.21
Nonane	26.67	712.34	714.24	713.13	4.82	0.63
Nonane	30.00	710.20	712.74	710.43	4.47	0.72
Nonane	32.22	707.54	710.90	706.49	4.57	1.22
Nonane	37.78	702.89	706.33	703.41	4.55	0.27
Nonane	40.00	702.50	705.08	700.24	4.59	0.95
Nonane	43.33	698.09	702.44	697.47	4.83	1.35
Nonane	48.89	693.44	697.27	691.93	5.13	1.40
Nonane	50.00	694.70	697.42	691.01	4.71	1.41
Nonane	54.44	688.64	693.97	686.63	4.97	2.15
Nonane	60.00	683.99	689.18	683.31	4.82	1.34
Nonane	65.56	679.18	685.49	676.77	5.10	2.66
Nonane	71.11	674.54	681.26	673.87	5.17	1.35

Nonane	76.67	669.73	676.99	669.15	5.01	1.36
Octadecane	50.00	763.20	772.89	763.86	6.48	0.86
Octadecane	75.00	744.60	753.70	747.01	6.80	0.47
Octadecane	100.00	726.20	734.85	726.05	7.41	0.86
Octadecane	125.00	709.20	716.90	710.71	8.11	0.63
Octadecane	150.00	691.80	696.74	689.94	8.01	1.77
Octadecane	175.00	674.20	676.40	674.67	8.01	0.70
Octadecane	200.00	656.20	659.64	656.64	9.14	0.96
Octane	25.00	698.60	708.01	700.27	4.40	0.64
Octane	50.00	678.20	688.99	677.93	4.59	0.88
Octane	75.00	657.20	669.89	655.41	4.79	2.22
Octane	100.00	635.30	650.89	635.14	5.21	2.12
Pentadecane	20.00	768.50	770.39	767.64	5.65	0.37
Pentadecane	30.00	761.60	763.01	761.23	5.93	0.18
Pentadecane	40.00	754.50	755.15	754.04	5.90	0.18
Pentadecane	50.00	747.40	748.24	747.69	6.43	0.83
Pentadecane	60.00	740.40	739.91	739.59	6.15	0.16
Pentadecane	70.00	733.30	732.38	732.24	6.14	0.20
Pentadecane	80.00	726.20	725.34	726.05	6.84	0.86
Pentadecane	90.00	719.10	716.68	717.31	6.37	0.39
Pentadecane	100.00	712.10	709.41	710.39	6.49	0.43
Tetradecane	10.00	770.00	769.85	770.04	5.47	0.85
Tetradecane	15.00	766.40	766.04	766.05	5.53	0.23
Tetradecane	20.00	762.90	762.03	762.47	5.45	0.23
Tetradecane	25.00	759.40	758.42	759.26	5.66	0.51
Tetradecane	35.00	752.30	750.60	751.75	5.63	0.06
Tetradecane	45.00	745.40	742.98	744.38	5.76	0.13
Tetradecane	65.00	731.40	727.54	729.80	5.87	0.15
Tetradecane	85.00	717.20	712.68	714.61	6.39	0.55
Tetradecane	100.00	706.50	700.52	703.81	6.30	0.77
Tridecane	20.00	755.80	753.87	757.01	5.58	0.83
Tridecane	30.00	748.70	746.20	748.87	5.51	0.12
Tridecane	40.00	741.60	738.60	741.71	5.39	0.14
Tridecane	40.00	741.90	738.60	741.71	5.39	0.14

Tridecane	50.00	734.70	730.81	734.42	5.61	0.29
Tridecane	60.00	727.50	723.34	727.47	6.14	0.66
Tridecane	70.00	720.50	715.57	719.28	5.86	0.55
Tridecane	80.00	715.00	707.52	712.52	5.86	0.16
Tridecane	90.00	705.70	699.80	705.12	5.98	0.18
Tridecane	100.00	698.50	692.08	697.94	6.10	0.28
Undecane	30.00	733.00	729.50	734.08	4.87	1.31
Undecane	50.00	718.50	713.70	719.88	5.57	0.43

Table A.8 Full density results. The table contains density experimental values, and linear regression and neural network cross-validation predictions accompanied by their uncertainties.

Name	$T(^{\circ}C)$	$\mu^{\exp}(cP)$	$\mu^{\mathrm{reg}}(cP)$	$\mu^{\mathrm{nn}}(cP)$	$\delta\mu^{\mathrm{reg}}(cP)$	$\delta\mu^{\mathrm{nn}}(cP)$
Octane	25.00	0.51	0.73	0.53	0.30	0.01
Octane	50.00	0.39	0.41	0.41	0.31	0.01
Octane	75.00	0.30	0.07	0.29	0.32	0.03
Octane	100.00	0.25	-0.28	0.22	0.36	0.02
Nonane	1.67	0.94	1.23	1.06	0.29	0.01
Nonane	4.44	0.90	1.20	0.95	0.30	0.03
Nonane	7.22	0.87	1.12	0.90	0.29	0.05
Nonane	10.00	0.83	1.12	0.85	0.30	0.02
Nonane	12.78	0.80	1.04	0.85	0.29	0.06
Nonane	15.56	0.77	1.03	0.70	0.29	0.03
Nonane	18.33	0.74	0.98	0.73	0.30	0.03
Nonane	21.11	0.71	0.93	0.69	0.30	0.02
Nonane	23.89	0.68	0.93	0.79	0.31	0.07
Nonane	26.67	0.66	0.88	0.77	0.31	0.08
Nonane	29.44	0.64	0.86	0.75	0.32	0.08
Nonane	30.00	0.62	0.84	0.63	0.31	0.01
Nonane	32.22	0.61	0.77	0.61	0.29	0.01
Nonane	35.00	0.59	0.79	0.59	0.32	0.01
Nonane	37.78	0.57	0.69	0.70	0.30	0.09
Nonane	40.00	0.55	0.71	0.56	0.32	0.01
Nonane	40.56	0.55	0.69	0.56	0.31	0.01

Nonane	43.33	0.54	0.65	0.55	0.32	0.01
Nonane	46.11	0.52	0.62	0.51	0.32	0.02
Nonane	48.89	0.50	0.53	0.49	0.31	0.02
Nonane	50.00	0.49	0.59	0.63	0.32	0.10
Decane	25.00	0.85	1.09	0.82	0.32	0.01
Decane	50.00	0.61	0.75	0.60	0.34	0.01
Decane	75.00	0.46	0.44	0.48	0.36	0.01
Decane	100.00	0.36	0.10	0.39	0.38	0.02
Undecane	30.00	0.99	1.22	0.98	0.33	0.05
Undecane	50.00	0.75	0.94	0.74	0.34	0.02
Dodecane	20.00	1.49	1.54	1.48	0.34	0.05
Dodecane	25.00	1.34	1.48	1.36	0.34	0.03
Dodecane	25.00	1.34	1.45	1.41	0.34	0.03
Dodecane	30.00	1.25	1.41	1.23	0.35	0.02
Dodecane	40.00	1.07	1.26	1.03	0.35	0.05
Dodecane	50.00	0.91	1.14	0.87	0.36	0.02
Dodecane	60.00	0.81	1.00	0.88	0.35	0.06
Dodecane	70.00	0.72	0.88	0.71	0.37	0.01
Dodecane	75.00	0.66	0.82	0.68	0.38	0.01
Dodecane	80.00	0.64	0.75	0.65	0.40	0.01
Dodecane	90.00	0.57	0.63	0.58	0.39	0.01
Dodecane	100.00	0.50	0.42	0.61	0.39	0.11
Dodecane	125.00	0.40	0.19	0.44	0.41	0.02
Dodecane	150.00	0.32	-0.12	0.32	0.43	0.01
Dodecane	175.00	0.26	-0.56	0.21	0.46	0.02
Dodecane	200.00	0.22	-1.02	0.09	0.46	0.03
Tridecane	20.00	1.80	1.76	1.95	0.36	0.02
Tridecane	30.00	1.57	1.62	1.51	0.37	0.06
Tridecane	40.00	1.29	1.50	1.20	0.35	0.02
Tridecane	50.00	1.10	1.33	1.04	0.37	0.03
Tridecane	60.00	0.97	1.19	0.89	0.38	0.03
Tridecane	70.00	0.85	1.06	0.82	0.39	0.01
Tridecane	80.00	0.75	0.92	0.75	0.40	0.01
Tridecane	90.00	0.67	0.81	0.68	0.42	0.02

Tridecane	100.00	0.59	0.63	0.63	0.40	0.01
Tetradecane	9.98	2.88	2.00	2.76	0.35	0.03
Tetradecane	12.17	2.74	2.04	2.67	0.36	0.02
Tetradecane	14.65	2.57	2.00	2.55	0.36	0.01
Tetradecane	14.65	2.58	1.94	2.55	0.36	0.02
Tetradecane	19.63	2.31	1.88	2.27	0.36	0.09
Tetradecane	25.58	2.04	1.80	2.12	0.37	0.02
Tetradecane	25.58	2.04	1.91	2.10	0.35	0.02
Tetradecane	44.82	1.41	1.56	1.35	0.38	0.03
Tetradecane	44.82	1.41	1.64	1.52	0.37	0.05
Tetradecane	44.83	1.41	1.56	1.47	0.38	0.04
Tetradecane	64.81	1.04	1.34	1.01	0.41	0.02
Tetradecane	64.83	1.04	1.31	1.01	0.40	0.02
Tetradecane	64.83	1.04	1.34	1.01	0.41	0.02
Tetradecane	64.83	1.04	1.34	1.01	0.41	0.02
Tetradecane	64.83	1.04	1.35	1.01	0.38	0.02
Tetradecane	84.46	0.80	1.07	0.89	0.41	0.06
Tetradecane	84.46	0.81	1.07	0.89	0.41	0.06
Tetradecane	84.46	0.81	1.07	0.89	0.41	0.06
Pentadecane	20.00	2.86	2.14	2.69	0.39	0.03
Pentadecane	30.00	2.29	1.93	2.30	0.39	0.02
Pentadecane	50.00	1.58	1.77	1.65	0.39	0.04
Pentadecane	60.00	1.35	1.56	1.33	0.41	0.04
Pentadecane	70.00	1.16	1.46	1.09	0.43	0.02
Pentadecane	80.00	1.01	1.31	1.02	0.43	0.03
Pentadecane	90.00	0.89	1.17	0.85	0.43	0.01
Pentadecane	100.00	0.79	1.05	0.78	0.42	0.01
Octadecane	50.00	2.46	2.24	2.53	0.44	0.02
Octadecane	75.00	1.60	1.96	1.66	0.46	0.03
Octadecane	100.00	1.12	1.60	1.11	0.48	0.03
Octadecane	125.00	0.83	1.30	0.79	0.49	0.02
Octadecane	150.00	0.65	0.94	0.70	0.52	0.03
Octadecane	175.00	0.52	0.68	0.57	0.53	0.02
Octadecane	200.00	0.42	0.26	0.46	0.56	0.01

Table A.9 Full dynamic viscosity results. The table contains density experimental values, and linear regression and neural network cross-validation predictions accompanied by their uncertainties.

Name	$\eta^{\exp}(cSt)$	$\mu^{\mathrm{fv}}(cSt)$	$\mu^{\mathrm{nn}}(cSt)$	$\delta\eta^{\mathrm{nn}}(cSt)$
Octane	0.65	0.72	0.77	0.04
Nonane	1.00	0.98	0.94	0.06
Decane	1.31	1.25	1.24	0.08
Undecane	1.56	1.77	1.68	0.12
Dodecane	1.96	1.97	2.19	0.13
Tridecane	2.48	3.12	2.55	0.14
Tetradecane	2.99	3.01	3.10	0.07
Pentadecane	3.78	5.50	3.57	0.09
Hexadecane	4.54	4.56	4.23	0.21

Table A.10 Predictions of kinematic viscosity of linear alkanes at 20° C. The table contains experimental values, free volume theory fits [68] and neural network predictions with their accompanying uncertainty.

Appendix B

Chapter 3: Full results

B.1 Density at 25°C

Name	$\rho_{exp}(\frac{g}{l})$	$\rho_{model}(\frac{g}{l})$	$\delta\rho_{model}(\frac{g}{l})$	$\lvert\Delta\rvert(\frac{g}{l})$
2,2-dimethylbutane	644.43	651.44	0.67	7.0
2,2-dimethylheptane	706.60	702.49	0.72	4.1
2,2-dimethylhexane	691.11	691.47	0.40	0.35
2,2-dimethyloctane	721.00	712.54	0.47	8.5
2,2-dimethylpentane	669.48	674.69	0.52	5.2
Decane	726.14	727.59	0.82	1.5
Heptane	679.50	678.40	0.42	1.1
Hexane	654.89	653.67	0.41	1.2
Nonane	713.75	714.01	0.97	0.26
Octane	698.76	699.37	0.49	0.61
3-ethylheptane	722.50	723.07	0.65	0.57
3-ethylhexane	709.45	709.08	0.77	0.37
3-ethyloctane	735.40	734.29	0.44	1.1
3-ethylpentane	693.92	690.89	0.52	3.0
3,3-diethylpentane	749.92	754.54	0.40	4.6
4-ethylheptane	722.30	722.38	0.37	0.08
4-ethyloctane	734.30	733.16	0.78	1.1
4-propylheptane	731.90	732.27	0.93	0.37
2-methylheptane	693.87	696.93	0.45	3.1
2-methylhexane	674.34	677.72	0.38	3.4

2-methylnonane	722.70	723.37	0.66	0.67
2-methyloctane	709.60	711.43	0.75	1.8
2-methylpentane	648.50	653.61	0.44	5.1
3-methylheptane	701.73	700.21	0.37	1.5
3-methylhexane	682.88	682.17	0.45	0.7
3-methylnonane	729.50	726.73	0.51	2.8
3-methyloctane	716.70	714.95	0.42	1.8
3-methylpentane	659.76	657.87	1.12	1.89
4-methylheptane	700.54	700.83	0.58	0.29
4-methylnonane	728.20	726.63	0.54	1.6
4-methyloctane	716.30	714.70	0.53	1.6
5-methylnonane	728.40	725.88	0.42	2.5
3-ethyl-2-methylheptane	739.80	729.11	0.43	11
3-ethyl-2-methylhexane	729.00	728.87	0.37	0.13
3-ethyl-2-methylpentane	715.20	717.26	0.51	2.1
3-ethyl-3-methylheptane	744.40	747.21	0.43	2.8
3-ethyl-3-methylhexane	736.00	736.93	0.53	0.93
3-ethyl-3-methylpentane	723.54	724.90	0.57	1.4
3-ethyl-4-methylheptane	746.60	742.77	0.31	3.8
3-ethyl-4-methylhexane	735.00	732.52	0.48	2.5
4-ethyl-2-methylheptane	732.20	734.77	0.37	2.6
4-ethyl-2-methylhexane	720.20	724.34	0.66	4.1
4-ethyl-3-methylheptane	746.80	742.22	0.33	4.6
4-ethyl-4-methylheptane	743.20	746.97	0.41	3.8
5-ethyl-2-methylheptane	731.50	733.73	0.40	2.2
5-ethyl-3-methylheptane	736.80	738.63	0.67	1.8
2,3-dimethylbutane	657.00	659.81	0.65	2.8
2,3-dimethylheptane	722.00	716.74	0.76	5.3
2,3-dimethylhexane	708.16	703.73	0.58	4.4
2,3-dimethyloctane	734.10	727.82	0.78	6.3
2,3-dimethylpentane	690.81	688.19	0.43	2.6
2,4-dimethylheptane	711.50	713.36	0.32	1.9
2,4-dimethylhexane	696.11	701.16	0.32	5.1
2,4-dimethyloctane	722.60	725.86	0.52	3.3

2,4-dimethylpentane	668.23	679.09	0.47	11
2,5-dimethylheptane	713.60	713.11	0.44	0.49
2,5-dimethylhexane	689.37	695.62	0.52	6.3
2,5-dimethyloctane	723.80	724.17	0.50	0.37
2,6-dimethylheptane	704.50	714.67	0.40	10.2
2,6-dimethyloctane	724.80	723.64	0.35	1.2
2,7-dimethyloctane	719.80	719.95	0.46	0.15
3,3-dimethylheptane	721.60	723.01	0.42	1.4
3,3-dimethylhexane	705.95	709.36	0.50	3.4
3,3-dimethyloctane	734.40	730.25	0.50	4.2
3,3-dimethylpentane	689.16	704.56	0.58	15
3,4-dimethylheptane	727.50	722.15	0.52	5.4
3,4-dimethylhexane	715.15	708.17	0.34	7.0
3,4-dimethyloctane	741.00	730.96	0.42	10
3,5-dimethylheptane	716.60	717.96	1.19	1.4
3,5-dimethyloctane	732.90	728.38	0.43	4.5
3,6-dimethyloctane	731.50	726.47	0.39	5.0
4,4-dimethylheptane	718.30	722.99	0.54	4.7
4,4-dimethyloctane	731.20	732.77	0.56	1.6
4,5-dimethyloctane	743.20	730.89	0.36	12

Table B.1 Results obtained from molecular dynamics simulations after correction factors in a leave-one-out cross validation are applied for density at 25°C. Experimental data is obtained from the TRC Thermodynamic tables [4].

B.2 Density at 100°C

Name	$\rho_{\text{exp}}(\frac{\text{g}}{\text{l}})$	$\rho_{\text{model}}(\frac{\text{g}}{\text{l}})$	$\delta\rho_{\text{model}}(\frac{\text{g}}{\text{l}})$	$\lvert\Delta\rvert(\frac{\text{g}}{\text{l}})$
Decane	667.70	667.99	0.32	0.29
Heptane	611.00	612.63	0.47	1.6
Hexane	581.40	579.66	0.85	1.7
Nonane	652.50	652.14	0.67	0.36
Octane	635.19	635.37	0.60	0.18

2-methylheptane	632.00	631.69	0.52	1.7
2-methylhexane	602.00	611.99	0.63	8.8
2-methylpentane	574.30	580.04	0.70	4.5
3-methylheptane	638.40	636.39	0.66	3.4
3-methylhexane	619.00	614.50	0.65	6.0
3-methylpentane	582.40	586.35	0.52	2.7
4-methylheptane	639.00	635.54	0.53	4.9
3-ethylhexane	647.00	644.06	0.66	2.9
3-ethylpentane	621.00	625.14	0.60	4.1
4-propylheptane	673.40	672.20	0.44	1.2
2,3-dimethylbutane	582.50	584.21	0.74	1.7
2,3-dimethylhexane	644.10	635.12	0.57	9.0
2,3-dimethylpentane	626.00	627.51	0.62	1.5
2,4-dimethylhexane	616.30	632.27	0.52	16
2,4-dimethylpentane	601.00	605.48	0.52	4.5
2,5-dimethylhexane	623.60	625.31	0.48	1.7
2,6-dimethylheptane	640.00	644.74	0.42	4.7
2,7-dimethyloctane	660.20	656.44	0.86	3.8
3,3-dimethylhexane	646.70	640.78	1.05	5.9
3,3-dimethylpentane	608.00	635.21	0.45	27
3,4-dimethylhexane	658.50	639.86	0.64	19
4,5-dimethyloctane	685.50	665.46	0.63	20
2,2-dimethylbutane	568.30	576.37	0.75	8.1
2,2-dimethylhexane	626.10	618.23	0.83	7.9
2,2-dimethylpentane	601.90	601.71	0.46	0.19
3-ethyl-2-methylpentane	657.00	638.35	0.45	19
3-ethyl-3-methylhexane	641.00	661.82	0.43	21
3-ethyl-3-methylpentane	663.30	648.87	0.85	14
5-ethyl-2-methylheptane	672.70	658.79	0.60	14

Table B.2 Results obtained from molecular dynamics simulations after correction factors in a leave-one-out cross validation are applied for density at 100°C. Experimental data is obtained from the TRC Thermodynamic tables [4].

B.3 Viscosity of linear alkanes

Name	$\rho_{exp}(\frac{g}{l})$	η_{exp} **(cSt)**	η_{pred} **(cSt)**	$\delta\eta_{\text{pred}}$ **(cSt)**	$\Delta_{\%}$
Hexane	659	0.46	0.43	0.04	-7
Heptane	684	0.60	0.68	0.08	13
Octane	703	0.78	0.78	0.04	0
Nonane	718	0.99	1.01	0.03	2
Decane	730	1.24	1.13	0.08	-9
Undecane	740	1.60	1.53	0.04	-4
Dodecane	749	2.00	1.81	0.12	-9.5
Tridecane	756	2.38	2.32	0.27	-2.5
Tetradecane	762	3.01	2.70	0.14	-10

Table B.3 Summary of viscosity simulations for linear alkanes at 20°. Alkane's name, experimental value of kinematic viscosity, simulation result, its uncertainty, and percent error are presented.

B.4 Viscosity of tridecane as a function of pressure at 60°C

p(MPa)	$\rho_{exp}(\frac{g}{l})$	η_{exp} **(cSt)**	η_{pred} **(cSt)**	$\delta\eta_{\text{pred}}$ **(cSt)**	$\Delta_{\%}$
0.1	728	1.33	1.38	0.16	3.8
20	743	1.67	1.67	0.06	0
40	757	2.02	2.09	0.18	3.5
60	768	2.36	2.23	0.08	-6
80	779	2.74	2.84	0.18	3.6
100	788	3.12	3.37	0.16	8.0

Table B.4 Results of viscosity simulations for tridecane at 60° as a function of pressure. Pressure, experimental value of kinematic viscosity, simulation result, its uncertainty, and percent error are presented.

B.5 Viscosity of octane, dodecane and octadecane as a function of temperature

T(°C)	$\rho_{exp}(\frac{g}{l})$	η_{exp} (cSt)	η_{pred} (cSt)	$\delta\eta_{\text{pred}}$ (cSt)	$\Delta_{\%}$
-10	729	1.15	1.10	0.05	-4
0	721	0.99	1.01	0.02	2
25	699	0.73	0.68	0.01	-7
40	686	0.63	0.66	0.03	5
60	669	0.53	0.50	0.05	-6
80	652	0.45	0.48	0.10	6.67
100	635	0.39	0.38	0.003	-3
125	618	0.33	0.34	0.02	3

Table B.5 Summary of viscosity simulations for octane as a function of temperature. Temperature, experimental value of kinematic viscosity, simulation result, its uncertainty, and percent error are presented.

T(°C)	$\rho_{exp}(\frac{g}{l})$	η_{exp} (cSt)	η_{pred} (cSt)	$\delta\eta_{\text{pred}}$ (cSt)	$\Delta_{\%}$
40	734	1.46	1.57	0.06	8
60	720	1.12	1.03	0.09	-8
80	704	0.90	0.87	0.04	-3
100	690	0.73	0.79	0.11	8.2
125	671	0.60	0.58	0.02	-3
150	651	0.50	0.54	0.03	8
175	630	0.42	0.47	0.03	12
200	609	0.36	0.40	0.02	11

Table B.6 Summary of viscosity simulations for dodecane as a function of temperature. Temperature, experimental value of kinematic viscosity, simulation result, its uncertainty, and percent error are presented.

T(°C)	$\rho_{exp}(\frac{g}{l})$	η_{exp} **(cSt)**	η_{pred} **(cSt)**	$\delta\eta_{\text{pred}}$ **(cSt)**	$\Delta_{\%}$
50	762	3.23	3.24	0.25	0.31
75	744	2.14	2.02	0.18	-5.7
100	727	1.55	1.65	0.20	6.5
125	709	1.18	1.12	0.12	-5.1
150	691	0.93	0.96	0.05	3
175	674	0.77	0.81	0.07	5
200	656	0.64	0.63	0.04	-2

Table B.7 Results of viscosity simulations for octadecane as a function of temperature. Temperature, experimental value of kinematic viscosity, simulation result, its uncertainty, and percent error are presented.

Appendix C

Chapter 4: Full results

C.1 Density at 25°C

Name	$\rho_{MD}(\frac{g}{l})$	$\rho_{ANN}(\frac{g}{l})$	$\delta\rho_{MD}(\frac{g}{l})$	$\delta\rho_{ANN}(\frac{g}{l})$
2-methyldecane	739.87	740.23	0.64	0.87
2,8-dimethylnonane	745.15	746.37	0.65	1.01
5-ethyl-2-methyloctane	750.72	751.31	0.50	0.69
2-methyl-4-propylheptane	750.49	751.21	0.58	1.95
2,5-dimethylnonane	748.06	748.77	0.48	0.37
3-ethyl-2-methyloctane	755.52	754.24	0.60	0.83
2,2-dimethylnonane	749.41	750.17	0.57	0.57
3-methyl-4-propylheptane	758.87	757.82	0.61	1.01
3,4-dimethylnonane	754.77	752.77	0.69	0.43
3-ethylnonane	745.91	744.96	0.50	0.74
3,5-diethylheptane	757.79	760.03	0.55	2.10
3-methyl-4-propylheptane	758.87	757.82	0.59	1.01
4-methyldecane	742.25	742.14	0.59	0.75
4-methyl-4-propylheptane	764.06	762.53	0.62	1.42
4,5-dimethylnonane	755.82	754.79	0.59	0.66
4,4-diethylheptane	771.22	769.66	0.60	0.98
4-propyloctane	744.74	744.51	0.50	0.54
5-ethylnonane	744.67	745.24	0.55	1.07
Dodecane	743.34	743.86	0.55	2.00
2-methylundecane	748.23	748.86	0.58	1.09

2,9-dimethyldecane	753	754.33	0.58	1.25
2,8-dimethyldecane	756.54	755.47	0.52	1.11
5-ethyl-2-methylnonane	758.03	759.20	0.60	0.61
2-methyl-4-propyloctane	757.56	759.13	0.56	0.96
2,5-dimethyldecane	754.26	755.95	0.60	0.81
2,4-dimethyldecane	754.93	756.44	0.58	0.70
3-ethyl-2-methylnonane	762.05	761.29	0.60	0.67
2,2-dimethyldecane	756.52	757.17	0.52	0.90
3-methylundecane	749.93	750.08	0.56	1.09
6-ethyl-3-methylnonane	760.62	761.09	0.68	1.86
3-methyl-5-propyloctane	760.68	762.00	0.66	1.06
3,5-dimethyldecane	758.67	758.71	0.62	0.50
3-ethyl-3-methylnonane	768.22	767.74	0.67	1.44
3,3-dimethyldecane	762.14	761.41	0.61	0.97
3-ethyldecane	753.96	752.64	0.54	0.54
3-ethyl-7-methynonane	764.7	763.71	0.53	1.14
3-ethyl-6-methylnonane	760.66	762.68	0.55	1.17
3,5-diethyloctane	765.15	767.26	0.58	1.91
3-ethyl-4-propylheptane	760.8	761.32	0.64	2.76
4,6-dimethyldecane	758.8	760.86	0.57	1.32
4-methyl-4-propyloctane	769.02	767.87	0.53	1.23
4,4-dimethyldecane	761.83	762.19	0.59	0.74
4-ethyldecane	751.18	752.93	0.51	0.59
4,5-diethyloctane	767.73	770.27	0.50	1.00
4-ethyl-4-propylheptane	774.79	771.38	0.58	2.45
4-propylnonane	751.45	752.38	0.57	0.72
5,6-dimethyldecane	760.88	762.77	0.44	0.66
5-ethyl-5-methylnonane	767.63	768.46	0.61	0.97
5-ethyldecane	753.34	752.64	0.68	0.88
5-propylnonane	752.47	752.27	0.61	0.27
6-methylundecane	749.4	750.12	0.54	1.43
2,4-dimethylundecane	761.39	762.81	0.62	0.69
7-ethyl-2-methylundecane	770.82	769.84	0.81	0.53
2-methyl-4-propyldecane	768.93	771.04	0.53	0.77

2,2-dimethyltridecane	773.28	772.20	0.57	2.27
9-ethyl-2-methyldodecane	773.84	773.77	0.62	1.44
4-ethyl-6-methyldodecane	777.4	779.53	0.7	0.98
5-ethyl-6-methyldodecane	780.24	781.61	0.73	2.04
2-methyl-4-propyldodecane	778.83	780.18	0.46	0.78
5-butyl-2-methylundecane	777.96	779.94	0.76	0.44
3-methyl-4-propyldodecane	782.86	782.64	0.61	1.03
7-ethyl-4-methyltridecane	779.92	781.16	0.72	0.80
4-methyl-4-propyldodecane	785.36	784.47	0.5	1.47
4-ethyl-8-methyltridecane	780.69	781.56	0.64	1.07
4,5-diethyldodecane	785.67	787.83	0.58	1.00
4,6-dipropyl-decane	781.42	782.15	0.57	1.81
6-ethyl-4-propylundecane	781.6	784.06	0.94	1.98
6-propyltridecane	773.47	774.19	0.62	0.26
3-ethyl-9-propyl-tridecane	789.52	789.51	0.87	1.67
4-methyl-6-propyl-tetradecane	788.01	789.34	0.53	1.35
6-butyl-6-methyl-tridecane	793.4	792.72	1.48	0.67
2,16-dimethyl-heptadecane	788.23	785.86	0.6	2.87
7-butyl-2-methyl-tetradecane	787.92	787.32	0.79	1.34
3,12-diethyl-pentadecane	793.22	792.68	0.62	0.74
4-methyl-7-pentyl-tridecane	788.57	788.18	0.81	1.11
11-ethyl-6-methyl-hexadecane	790.07	790.98	0.95	1.21
6,11-dimethyl-heptadecane	787.77	790.41	0.58	1.54
7-ethyl-8-propyl-tetradecane	799.09	797.18	1.19	0.70
8-ethyl-heptadecane	784.09	784.27	0.61	1.02
8,8-diethyl-pentadecane	799.93	798.26	0.84	0.98
2,11-dimethyl-octadecane	791.21	789.53	0.5	2.15
3-ethyl-octadecane	788.98	786.37	0.59	1.32
4-ethyl-8-methyl-heptadecane	792.57	792.88	1.44	2.01
6-ethyl-10-propyl-pentadecane	794.95	796.67	0.91	0.89
6,6-diethyl-hexadecane	801.93	798.20	0.63	1.09
7-methyl-nonadecane	787.5	786.52	0.85	1.78
6,12-dimethyl-octadecane	796.62	792.50	0.69	1.49
9-ethyl-7-methyl-heptadecane	791.97	794.77	0.68	0.82

7-butyl-7-ethyl-tetradecane	800.67	799.19	1.48	1.84
8,8-dimethyl-octadecane	793.21	792.91	0.89	0.66
8-ethyl-octadecane	787.91	787.11	0.48	0.67
8,8-diethyl-hexadecane	802.36	799.92	2.56	1.20
7-methyl-9-pentyl-pentadecane	793.16	791.18	0.69	1.73
4-methyl-9-propyl-heptadecane	795.2	795.15	1.03	1.50
4-ethyl-11-propyl-hexadecane	795.8	796.72	0.84	1.39
5,6-diethyl-heptadecane	799.24	799.25	0.8	0.79
7,8-diethyl-heptadecane	797.51	800.13	0.5	0.84
11-methyl-4-propyl-octadecane	794.59	796.75	0.8	3.23
7-pentyl-6-methyl-pentadecane	795.78	797.60	0.91	0.96
7,8-diethyl-octadecane	800.81	802.12	1.08	0.50
8-heptyl-pentadecane	791.57	791.02	0.88	0.65
10-hexyl-5-methyl-hexadecane	797.65	797.35	0.54	0.67
7-hexyl-heptadecane	794.34	793.78	1	0.59
8-ethyl-9-propyl-octadecane	800.88	804.49	0.56	1.09
8-heptyl-hexadecane	794.23	793.19	1.03	0.29
7-ethyl-6-pentyl-heptadecane	801.23	801.66	1.45	1.56
7-hexyl-octadecane	795.36	795.63	0.82	0.54
8-heptyl-heptadecane	795.72	795.09	1.33	0.36
9-heptyl-heptadecane	795.46	795.19	1.55	0.40
2,8-dimethyl-tricosane	799.19	798.79	1.1	1.35
3-ethyl-tricosane	791.77	795.86	0.54	2.81
3,19-diethyl-henicosane	804.07	803.02	0.73	1.57
3-ethyl-17-propyl-icosane	805.42	803.66	1.03	1.31
15-butyl-3-ethyl-nonadecane	804.69	802.31	0.77	2.28
3,5-diethyl-henicosane	804.79	803.53	0.68	0.45
4-ethyl-9-propyl-icosane	804.41	804.54	1.21	1.33
13-butyl-6-propyl-octadecane	799.76	800.59	1.35	1.21
7-hexyl-nonadecane	799.38	797.34	1.35	0.57
8-heptyl-octadecane	795.15	796.77	0.94	0.54
9,9-dipropyl-nonadecane	807.2	805.95	0.5	2.21
9-heptyl-octadecane	796.32	796.94	0.67	0.25
9-octyl-heptadecane	797.11	796.59	0.93	0.85

9-butyl-4-ethyl-tricosane	810.16	809.09	1.27	1.21
6-butyl-8-propyl-pentacosane	811.24	810.93	1.32	1.21
7,22-dimethyl-tricosane	808.58	808.40	1.07	1.39
13-butyl-8-propyl-pentacosane	809.01	810.50	1.15	1.65
14-butyl-5-methyl-nonacosane	812.51	812.54	1.01	1.27
8-heptyl-octacosane	806.61	807.88	1.03	0.99
7-butyl-4-ethyl-hentriacontane	814.1	814.92	0.93	0.81
15-butyl-16-methyl-tritriacontane	814.33	814.31	1.5	0.33
16-butyl-16-methyl-tritriacontane	814.63	814.38	1.17	0.36
18-nonyl-5-propyl-heptacosane	812.56	813.37	1.35	0.87
15-butyl-16-methyl-tetratriacontane	815.5	814.84	0.81	0.44
16-butyl-17-methyl-tetratriacontane	813.46	814.86	0.97	0.41
10,20-dipentyl-tricosane	817.21	816.70	2.02	0.90
15-butyl-17-methyl-pentatriacontane	815.27	815.34	1.27	0.46
15-butyl-16-methyl-pentatriacontane	815.13	815.34	0.73	0.57
16-butyl-19-methyl-pentatriacontane	814.02	815.37	1.04	0.40
16-butyl-16-methyl-pentatriacontane	815.81	815.35	0.81	0.36
16-butyl-17-methyl-pentatriacontane	816.72	815.34	1.6	0.47
17-butyl-18-methyl-pentatriacontane	816.51	815.52	0.83	0.66

Table C.1 MD and ANN results for density at 25°C.

C.2 Density at 100°C

Name	$\rho_{\mathrm{MD}}(\frac{\mathrm{g}}{\mathrm{l}})$	$\rho_{\mathrm{ANN}}(\frac{\mathrm{g}}{\mathrm{l}})$	$\delta\rho_{\mathrm{MD}}(\frac{\mathrm{g}}{\mathrm{l}})$	$\delta\rho_{\mathrm{ANN}}(\frac{\mathrm{g}}{\mathrm{l}})$
2-methyldecane	683.56	683.37	0.56	0.63
2,8-dimethylnonane	687.97	691.15	0.58	1.10
5-ethyl-2-methyloctane	693.84	694.52	0.54	0.69
2-methyl-4-propylheptane	693.52	692.38	0.53	1.30
2,5-dimethylnonane	690.54	692.92	0.53	0.92
3-ethyl-2-methyloctane	700.13	698.44	0.58	0.92
2,2-dimethylnonane	693.85	694.84	0.52	0.64
3-methyl-4-propylheptane	703.49	702.10	0.61	1.00

3,4-dimethylnonane	699.78	697.01	0.61	0.86
3-ethylnonane	690.16	688.82	0.53	0.32
3,5-diethylheptane	703.15	704.43	0.53	0.90
3-methyl-4-propylheptane	703.49	702.10	0.61	1.00
4-methyldecane	685.28	685.21	0.56	0.17
4-methyl-4-propylheptane	707.3	707.33	0.63	0.92
4,5-dimethylnonane	698.33	699.25	0.56	0.71
4,4-diethylheptane	716.48	714.42	0.58	1.61
4-propyloctane	688.54	688.44	0.55	0.61
5-ethylnonane	687.83	687.70	0.52	0.84
Dodecane	687.07	687.64	0.52	0.74
2-methylundecane	692.99	693.08	0.57	0.53
2,9-dimethyldecane	697.67	699.75	0.55	0.67
2,8-dimethyldecane	701.56	699.83	0.57	1.11
5-ethyl-2-methylnonane	702.77	703.52	0.59	0.67
2-methyl-4-propylheptane	693.52	692.38	0.53	1.30
2,5-dimethyldecane	699.37	701.39	0.57	0.98
2,4-dimethyldecane	700.86	701.86	0.60	0.95
3-ethyl-2-methylnonane	707.47	705.51	0.55	1.29
2,2-dimethyldecane	702.39	703.30	0.60	1.10
3-methylundecane	694.95	694.62	0.56	0.70
6-ethyl-3-methylnonane	705.09	706.12	0.65	0.41
3-methyl-5-propyloctane	705.8	708.19	0.57	1.08
3,5-dimethyldecane	703.32	704.13	0.58	0.57
3-ethyl-3-methylnonane	714.53	710.36	0.61	1.46
3,3-dimethyldecane	709.74	708.49	0.55	0.82
3-ethyldecane	698.7	698.47	0.53	0.37
3-ethyl-7-methynonane	710.16	707.84	0.58	0.89
3-ethyl-6-methylnonane	707.06	707.64	0.54	1.84
3,5-diethyloctane	709.52	711.46	0.62	1.10
3-ethyl-4-propylheptane	705.96	707.47	0.57	1.52
4,6-dimethyldecane	703.9	706.01	0.56	0.38
4-methyl-4-propyloctane	714.69	713.18	0.55	1.38
4,4-dimethyldecane	707.34	706.78	0.61	0.81

4-ethyldecane	696.12	698.12	0.52	0.98
4,5-diethyloctane	712.57	713.22	0.53	1.02
4-ethyl-4-propylheptane	722.1	719.18	0.62	1.49
4-propylnonane	697.59	698.13	0.57	0.75
5,6-dimethyldecane	707.44	707.83	0.53	1.18
5-ethyl-5-methylnonane	713.44	714.79	0.64	1.65
5-ethyldecane	697.23	697.23	0.61	1.01
5-propylnonane	697.29	697.78	0.62	0.46
6-methylundecane	693.73	693.60	0.53	1.72
2,4-dimethylundecane	708.16	709.23	0.64	0.99
7-ethyl-2-methylundecane	716.98	716.45	0.72	1.15
2-methyl-4-propyldecane	715.94	716.09	0.58	0.57
2,2-dimethyltridecane	720.26	719.42	0.72	1.18
9-ethyl-2-methyldodecane	722.76	721.45	0.48	1.09
4-ethyl-6-methyldodecane	724.75	727.07	0.68	0.94
5-ethyl-6-methyldodecane	727.39	727.89	0.51	0.89
2-methyl-4-propyldodecane	726.97	726.92	0.45	0.82
5-butyl-2-methylundecane	725.35	727.42	0.59	0.86
3-methyl-4-propyldodecane	731.22	729.86	0.65	0.76
7-ethyl-4-methyltridecane	727.99	730.05	0.7	1.50
4-methyl-4-propyldodecane	735.94	734.73	0.69	1.20
4-ethyl-8-methyltridecane	728.5	730.64	0.79	0.87
4,5-diethyldodecane	734.46	733.74	0.77	1.03
4,6-dipropyl-decane	728.62	731.38	0.59	1.91
6-ethyl-4-propylundecane	728.86	731.22	0.92	1.00
6-propyltridecane	722.19	724.39	0.77	1.47
3-ethyl-9-propyl-tridecane	738.22	737.79	0.82	0.65
4-methyl-6-propyl-tetradecane	733.43	737.41	0.92	0.71
6-butyl-6-methyl-tridecane	740.05	739.04	0.79	0.41
2,16-dimethyl-heptadecane	737.69	735.40	0.83	1.48
7-butyl-2-methyl-tetradecane	736.95	736.22	1.09	1.00
3,12-diethyl-pentadecane	743.37	740.98	0.57	0.86
4-methyl-7-pentyl-tridecane	736.53	738.50	0.62	0.73
11-ethyl-6-methyl-hexadecane	740.28	741.06	0.68	2.28

6,11-dimethyl-heptadecane	737.67	741.75	0.44	0.76
7-ethyl-8-propyl-tetradecane	751.21	746.68	0.79	1.69
8-ethyl-heptadecane	735.92	735.29	0.66	0.34
8,8-diethyl-pentadecane	748.29	748.07	0.49	1.09
2,11-dimethyl-octadecane	739.84	739.45	0.66	1.26
3-ethyl-octadecane	740.07	738.30	0.66	0.45
4-ethyl-8-methyl-heptadecane	744.43	745.25	1.06	1.69
6-ethyl-10-propyl-pentadecane	742.23	746.65	0.52	1.72
6,6-diethyl-hexadecane	752.01	749.21	0.85	1.18
7-methyl-nonadecane	737.54	737.97	0.63	2.48
6,12-dimethyl-octadecane	748.63	745.21	0.43	1.78
9-ethyl-7-methyl-heptadecane	742.12	744.21	0.48	2.13
7-butyl-7-ethyl-tetradecane	749.98	749.19	0.64	2.37
8,8-dimethyl-octadecane	744.51	745.54	0.63	1.40
8-ethyl-octadecane	738.42	738.36	0.54	0.58
8,8-diethyl-hexadecane	752.21	749.55	0.33	0.70
7-methyl-9-pentyl-pentadecane	743.09	740.46	0.67	1.01
4-methyl-9-propyl-heptadecane	745.21	744.84	0.55	0.76
4-ethyl-11-propyl-hexadecane	747.82	747.14	0.56	0.89
5,6-diethyl-heptadecane	749.22	749.70	0.41	1.48
7,8-diethyl-heptadecane	749.1	750.67	0.42	0.95
11-methyl-4-propyl-octadecane	747.29	747.19	0.7	1.54
7-pentyl-6-methyl-pentadecane	746.56	747.92	0.58	1.40
7,8-diethyl-octadecane	751.43	752.30	0.41	0.66
8-heptyl-pentadecane	741.16	741.74	0.46	0.99
10-hexyl-5-methyl-hexadecane	748.26	748.72	0.57	1.03
7-hexyl-heptadecane	745.26	745.41	0.59	0.40
8-ethyl-9-propyl-octadecane	751.87	753.95	2.56	1.51
8-heptyl-hexadecane	744.81	744.54	0.43	0.47
7-ethyl-6-pentyl-heptadecane	753.49	753.45	0.5	1.52
7-hexyl-octadecane	747.57	747.78	0.59	0.41
8-heptyl-heptadecane	746.41	747.07	0.71	0.21
9-heptyl-heptadecane	747.02	747.39	0.66	0.54
2,8-dimethyl-tricosane	752.92	754.19	0.64	1.91

3-ethyl-tricosane	745.07	747.80	0.44	1.66
3,19-diethyl-henicosane	756.16	755.10	0.51	1.21
3-ethyl-17-propyl-icosane	757.84	755.54	0.66	1.33
15-butyl-3-ethyl-nonadecane	755.97	754.53	0.54	1.10
3,5-diethyl-henicosane	757	756.76	0.73	0.73
4-ethyl-9-propyl-icosane	755.2	755.51	0.64	0.48
13-butyl-6-propyl-octadecane	753.76	755.47	0.73	2.24
7-hexyl-nonadecane	750.63	749.91	0.53	0.48
8-heptyl-octadecane	749.65	749.37	0.7	0.42
9,9-dipropyl-nonadecane	759.87	760.17	0.57	3.69
9-heptyl-octadecane	750.24	749.84	0.64	1.05
9-octyl-heptadecane	748.94	748.63	0.57	0.50
9-butyl-4-ethyl-tricosane	761.94	761.49	0.33	1.34
6-butyl-8-propyl-pentacosane	765.92	765.62	0.68	1.12
7,22-dimethyl-tricosane	763.32	763.97	0.51	0.97
13-butyl-8-propyl-pentacosane	765.59	766.15	0.49	1.04
14-butyl-5-methyl-nonacosane	767.87	766.00	2.12	1.62
8-heptyl-octacosane	764.34	764.26	0.52	1.63
7-butyl-4-ethyl-hentriacontane	770.72	770.83	1.18	0.50
15-butyl-16-methyl-tritriacontane	771.31	771.10	1.76	0.78
16-butyl-16-methyl-tritriacontane	770.66	770.81	0.52	0.55
18-nonyl-5-propyl-heptacosane	764.52	766.24	3.21	1.31
15-butyl-16-methyl-tetratriacontane	772.77	772.01	0.61	0.56
16-butyl-17-methyl-tetratriacontane	771.15	771.78	0.62	0.32
10,20-dipentyl-tricosane	773.02	771.97	2.43	2.26
15-butyl-17-methyl-pentatriacontane	770.68	772.74	0.64	0.48
15-butyl-16-methyl-pentatriacontane	773.44	772.92	1.11	0.46
16-butyl-19-methyl-pentatriacontane	772.13	772.50	0.63	0.53
16-butyl-16-methyl-pentatriacontane	772.75	772.60	0.53	0.40
16-butyl-17-methyl-pentatriacontane	773.6	772.74	1.15	0.31
17-butyl-18-methyl-pentatriacontane	773.08	772.64	0.56	0.54

Table C.2 MD and ANN results for density at 100°C.

C.3 Viscosity simulation results

$\dot{\gamma}$	$\eta(\dot{\gamma})$ (**cSt**)	$\delta\eta(\dot{\gamma})$ (**cSt**)
9.06	15.92	0.44
8.84	20.25	0.73
8.96	16.36	0.55
8.66	19.26	1.13
9.14	13.72	0.39
9.61	8.22	0.17
10.09	4.08	0.07
10.57	1.98	0.04
11.05	1.01	0.01
11.52	0.60	0.00
12.00	0.46	0.00

Table C.3 Simulation results for 8,11-dipentyloctadecane at 40°C.

$\dot{\gamma}$	$\eta(\dot{\gamma})$ (**cSt**)	$\delta\eta(\dot{\gamma})$ (**cSt**)
10.01	2.72	0.06
9.79	3.14	0.10
9.91	2.95	0.07
9.61	3.56	0.14
10.09	2.40	0.05
10.57	1.38	0.04
11.05	0.82	0.01
11.52	0.52	0.00
12.00	0.42	0.00

Table C.4 Simulation results for 8,11-dipentyloctadecane at 100°C

$\dot{\gamma}$	$\eta(\dot{\gamma})$ (**cSt**)	$\delta\eta(\dot{\gamma})$ (**cSt**)
9.06	18.91	0.40
8.84	22.78	0.66
8.96	18.88	0.48
8.66	24.92	0.98
9.14	17.93	0.33
9.61	9.13	0.15
10.09	4.27	0.05
10.57	2.03	0.03
11.05	1.01	0.01
11.52	0.61	0.00
12.00	0.46	0.00

Table C.5 Simulation results for 8,14-dipentylhenicosane at 40°C

$\dot{\gamma}$	$\eta(\dot{\gamma})$ (**cSt**)	$\delta\eta(\dot{\gamma})$ (**cSt**)
9.53	4.51	0.17
9.31	5.00	0.20
9.44	4.13	0.22
9.14	4.82	0.33
9.61	4.17	0.16
10.09	2.69	0.05
10.57	1.52	0.03
11.05	0.83	0.01
11.52	0.53	0.00
12.00	0.42	0.00

Table C.6 Simulation results for 8,14-dipentylhenicosane at 100°C.

$\dot{\gamma}$	$\eta(\dot{\gamma})$ (**cSt**)	$\delta\eta(\dot{\gamma})$ (**cSt**)
9.06	12.76	0.38
8.84	14.74	0.62
8.96	13.78	0.48
8.66	14.22	0.96
9.14	11.61	0.32
9.61	6.55	0.14
10.09	3.47	0.05
10.57	1.71	0.03
11.05	0.91	0.01
11.52	0.54	0.00
12.00	0.42	0.00

Table C.7 Simulation results for 11-heptyltricosane at 40°C.

$\dot{\gamma}$	$\eta(\dot{\gamma})$ (**cSt**)	$\delta\eta(\dot{\gamma})$ (**cSt**)
10.01	2.38	0.06
9.79	2.59	0.09
9.91	2.51	0.07
9.61	3.09	0.14
10.09	2.12	0.05
10.57	1.23	0.03
11.05	0.74	0.01
11.52	0.47	0.00
12.00	0.39	0.00

Table C.8 Simulation results for 11-heptyltricosane at 100°C.

Appendix D

Chapter 5: Full results

D.1 Model fits

x_{nonane}	$V_{\text{exp}}(\frac{\text{cm}^3}{\text{mol}})$	$V_{\text{fit}}(\frac{\text{cm}^3}{\text{mol}})$
1	179.61±0.01	179.62±0.01
0.92282	178.52±0.01	178.54±0.01
0.78722	176.64±0.01	176.64±0.01
0.6706	175.03±0.01	175.04±0.01
0.59027	173.94±0.01	173.94±0.01
0.47483	172.39±0.01	172.39±0.01
0.30417	170.10±0.01	170.10±0.01
0.28569	169.86±0.01	169.86±0.01
0.18143	168.48±0.01	168.48±0.01
0.11767	167.65±0.01	167.65±0.01
0.07895	167.15±0.01	167.15±0.01
0.0339	166.56±0.01	166.56±0.01

Table D.1 Molar volume of the mixture of nonane and isooctane at 25°C. Experimental data was obtained from Awwad and Allos[10].

x_{dodecane}	$V_{\text{exp}}(\frac{\text{cm}^3}{\text{mol}})$	$V_{\text{fit}}(\frac{\text{cm}^3}{\text{mol}})$
1	228.57±0.01	228.60±0.01
0.9077	222.72±0.01	222.71±0.01
0.84311	218.63±0.01	218.61±0.01
0.74973	212.72±0.01	212.70±0.01
0.62534	204.87±0.01	204.86±0.01
0.52136	198.35±0.01	198.34±0.01
0.4111	191.45±0.01	191.46±0.01
0.32383	186.04±0.01	186.04±0.01
0.24026	180.84±0.01	175.55±0.01
0.1541	175.52±0.01	175.55±0.01
0.09376	171.84±0.01	171.84±0.01
0.05433	169.42±0.01	169.42±0.01
0.02669	167.76±0.01	167.72±0.01

Table D.2 Molar volume of the mixture of dodecane and isooctane at 25°C. Experimental data was obtained from Awwad and Allos[10].

x_{heptane}	$\sigma_{\text{exp}}(\frac{\text{mN}}{\text{m}})$	$\sigma_{\text{fit}}(\frac{\text{mN}}{\text{m}})$
0	18.15±0.03	18.16
0.1	18.05±0.03	18.05
0.2	17.95±0.03	17.95
0.3	17.85±0.03	17.85
0.4	17.75±0.03	17.75
0.5	17.65±0.03	17.65
0.6	17.56±0.03	17.56
0.7	17.47±0.03	17.47
0.8	17.39±0.03	17.39
0.9	17.30±0.03	17.30
1	17.23±0.03	17.23

Table D.3 Surface tension of the mixture of heptane and isooctane at 40°C. Experimental data was obtained from S. López-Lázaro et al. [34]

x_{octane}	$\sigma_{\text{exp}}(\frac{\text{mN}}{\text{m}})$	$\sigma_{\text{fit}}(\frac{\text{mN}}{\text{m}})$
0	19.82±0.03	19.82
0.1	19.55±0.03	19.55
0.2	19.28±0.03	19.27
0.3	19.01±0.03	19.01
0.4	18.74±0.03	18.74
0.5	18.48±0.03	18.48
0.6	18.22±0.03	18.23
0.7	17.97±0.03	17.97
0.8	17.72±0.03	17.72
0.9	17.47±0.03	17.47
1	17.23±0.03	17.23

Table D.4 Surface tension of the mixture of octane and isooctane at 40°C. Experimental data was obtained from S. López-Lázaro et al. [34]

x_{nonane}	$K_{\text{S,m}-\text{exp}}(\text{m}^3(\text{TPa})^{-1}\text{mol}^{-1})$	$K_{\text{S,m}-\text{fit}}(\text{m}^3(\text{TPa})^{-1}\text{mol}^{-1})$
0	0.06391 ± 0.00013	0.06385 ± 0.00003
0.096	0.07257 ± 0.00015	0.07260 ± 0.00003
0.1942	0.08141 ± 0.00016	0.08148 ± 0.00003
0.2952	0.09054 ± 0.00018	0.09054 ± 0.00003
0.3954	0.09946 ± 0.00020	0.09946 ± 0.00003
0.4954	0.1083 ± 0.00022	0.10828 ± 0.00003
0.5929	0.11683 ± 0.00023	0.11682 ± 0.00004
0.6939	0.12562 ± 0.00025	0.12559 ± 0.00004
0.7939	0.13422 ± 0.00027	0.13420 ± 0.00004
0.8947	0.14279 ± 0.00029	0.14281 ± 0.00004
1	0.1517 ± 0.00030	0.15172 ± 0.00005

Table D.5 Molar isentropic compressibility of mixture of nonane and toluene at 10°C. Experimental data was obtained from González et al [42].

x_{hexane}	$K_{\text{S,m}-\text{exp}}(\text{m}^3(\text{TPa})^{-1}\text{mol}^{-1})$	$K_{\text{S,m}-\text{fit}}(\text{m}^3(\text{TPa})^{-1}\text{mol}^{-1})$
0	0.0639 ± 0.00013	0.06390 ± 0.00006
0.083	0.06931 ± 0.00014	0.06932 ± 0.00007
0.1713	0.07535 ± 0.00015	0.07537 ± 0.00007
0.298	0.08463 ± 0.00017	0.08459 ± 0.00008
0.396	0.09217 ± 0.00018	0.09215 ± 0.00008
0.5011	0.10076 ± 0.00020	0.10067 ± 0.00008
0.5949	0.10865 ± 0.00022	0.10864 ± 0.00008
0.6973	0.11781 ± 0.00024	0.11774 ± 0.00008
0.7955	0.12674 ± 0.00025	0.12684 ± 0.00007
0.8949	0.13637 ± 0.00027	0.13643 ± 0.00007
1	0.14717 ± 0.00029	0.14700 ± 0.00006

Table D.6 Molar isentropic compressibility of mixture of hexane and toluene at 10°C. Experimental data was obtained from González et al [42].

x_{heptane}	x_{octane}	x_{undecane}	$\mu_{\text{exp}}(\text{cP})$	$\mu_{\text{fit}}(\text{cP})$
0.1641	0.1442	0.0999	1.1388±0.0011	1.1379±0.0338
0.1831	0.1419	0.3025	1.0135±0.0010	1.0128±0.0222
0.1767	0.1413	0.5794	0.8855±0.0009	1.8857±0.0338
0.1676	0.3678	0.1004	0.8724±0.0009	0.8724±0.0139
0.1528	0.3674	0.3159	0.8002±0.0008	0.7995±0.0220
0.1679	0.5003	0.1067	0.7364±0.0007	0.7339±0.0159
0.4303	0.1366	0.0948	0.7952±0.0008	0.7935±0.0172
0.4293	0.1376	0.2687	0.7241±0.0007	0.7221±0.0200
0.3628	0.3421	0.1008	0.6699±0.0007	0.6692±0.0197
0.3264	0.1127	0.0787	0.9620±0.0010	0.9579±0.0292
0.0554	0.0316	0.1948	1.4098±0.0014	1.4088±0.0415
0.067	0.0665	0.7109	1.0535±0.0011	1.0529±0.0120
0.1107	0.656	0.0913	0.6570±0.0007	1.6575±0.0139

Table D.7 Viscosity of the mixture of heptane, octane, undecane and tridecane at 25°C. Experimental data was obtained from Wu et al [126].

Name	x_{heavier}	$\mu_{\text{exp}}(cP)$
tetradecane,hexane	0.8002	1.5462
tetradecane,hexane	0.4999	0.9402
tetradecane,hexane	0.3351	0.6820
hexadecane,hexane	0.7988	2.1936
hexadecane,hexane	0.5013	1.2372
hexadecane,hexane	0.2002	0.5844
hexadecane,tetradecane	0.8002	2.8105
hexadecane,tetradecane	0.4998	2.5085
hexadecane,tetradecane	0.2068	2.2307

Table D.8 Viscosity of the binary mixture of hexane and tetradecane, hexane and hexadecane, and tetradecane and hexadecane at 25 °C. Theory fits the data perfectly. Experimental data comes from Chevalier et al [22].

D.2 Viscosity predictions

$x_{\text{hexadecane}}$	$x_{\text{tetradecane}}$	μ_{exp}(cP)	μ_{pred}(cP)
0.065	0.0796	0.4631±0.0005	0.4712
0.1028	0.1119	0.5654±0.0006	0.5649
0.1046	0.1763	0.6538±0.0007	0.6518
0.1521	0.1151	0.6582±0.0007	0.6524
0.165	0.1831	0.7862±0.0008	0.7759
0.0878	0.3651	0.9111±0.0009	0.9069
0.2118	0.2258	0.9491±0.0010	0.9463
0.3274	0.0947	0.9851±0.0010	0.9814
0.1297	0.4208	1.1144±0.0011	1.1060
0.2648	0.2924	1.2078±0.0012	1.2018
0.3995	0.1514	1.2787±0.0013	1.2748
0.1105	0.5396	1.3054±0.0013	1.2931
0.2522	0.3998	1.4001±0.0014	1.3928
0.3777	0.2696	1.4749±0.0015	1.4701
0.5182	0.1243	1.5625±0.0016	1.5605

Table D.9 Viscosity of the ternary mixture of hexane, tetradecane, and hexadecane at 25 °C. Experimental data comes from Heric and Brewer [47].

Mixture	x_1	μ_{exp}(cP)	μ_{pred}(cP)
octane,heptane	0.9008	0.4940±0.0005	0.4964±0.0096
octane,heptane	0.5999	0.4558±0.0005	0.4575±0.0173
octane,heptane	0.3027	0.4216±0.0004	0.4219±0.0248
octane,heptane	0.7713	0.4777±0.0005	0.4793±0.0129
octane,heptane	0.4948	0.4442±0.0004	0.4446±0.2000
octane,heptane	0.2102	0.4110±0.0004	0.4114±0.0271
octane,heptane	0.7076	0.4701±0.0005	0.4711±0.0145
octane,heptane	0.3918	0.4320±0.0004	0.4323±0.0226
octane,heptane	0.1046	0.3990±0.0004	0.3998±0.0297
octane,undecane	0	1.081±0.0011	1.0640±0.0060
octane,undecane	0.0992	1.012±0.0010	1.0035±0.0078
octane,undecane	0.2059	0.9412±0.0009	0.9399±0.0094

octane,undecane	0.2978	0.8829±0.0009	0.8861±0.0104
octane,undecane	0.3916	0.8255±0.0008	0.8324±0.0111
octane,undecane	0.5019	0.7644±0.0008	0.7705±0.0115
octane,undecane	0.5959	0.7093±0.0007	0.7190±0.0115
octane,undecane	0.7154	0.6474±0.0006	0.6551±0.0111
octane,undecane	0.8041	0.6023±0.0006	0.6088±0.0104
octane,undecane	0.9011	0.5559±0.0006	0.5592±0.0094
octane,undecane	1	0.5105±0.0005	0.5099±0.0080
octane,tridecane	0	1.688±0.0017	1.6680±0.0080
octane,tridecane	0.1246	1.492±0.0015	1.4787±0.0095
octane,tridecane	0.1988	1.383±0.0014	1.3720±0.0102
octane,tridecane	0.3028	1.239±0.0012	1.2302±0.0110
octane,tridecane	0.3936	1.112±0.0011	1.1137±0.0113
octane,tridecane	0.4971	0.9963±0.0010	0.9891±0.0115
octane,tridecane	0.5956	0.885±0.0009	0.8788±0.0114
octane,tridecane	0.697	0.7811±0.0008	0.7736±0.0110
octane,tridecane	0.7898	0.6921±0.0007	0.6848±0.0103
octane,tridecane	0.8988	0.5929±0.0006	0.5896±0.0093
octane,tridecane	1	0.5105±0.0005	0.5099±0.0080
undecane,tridecane	0	1.688±0.0017	1.6680±0.0080
undecane,tridecane	0.1023	1.619±0.0016	1.6058±0.0089
undecane,tridecane	0.2031	1.549±0.0015	1.5447±0.0095
undecane,tridecane	0.2991	1.487±0.0015	1.4865±0.0099
undecane,tridecane	0.3981	1.423±0.0014	1.4266±0.0101
undecane,tridecane	0.4981	1.362±0.0014	1.3661±0.0100
undecane,tridecane	0.5976	1.304±0.0013	1.3061±0.0097
undecane,tridecane	0.7028	1.244±0.0012	1.2427±0.0091
undecane,tridecane	0.8	1.189±0.0012	1.1842±0.0083
undecane,tridecane	0.9004	1.134±0.0013	1.1238 ± 0.0073
undecane,tridecane	1	1.081±0.0011	1.0640±0.0060

Table D.10 Viscosity of mixtures of octane with heptane, undecane, or tridecane, and the mixture of undecane and tridecane. Experimental data for the mixture of heptane and octane has been obtained from Chevalier et al. [22], with remaining data coming from Wu et al. [61].