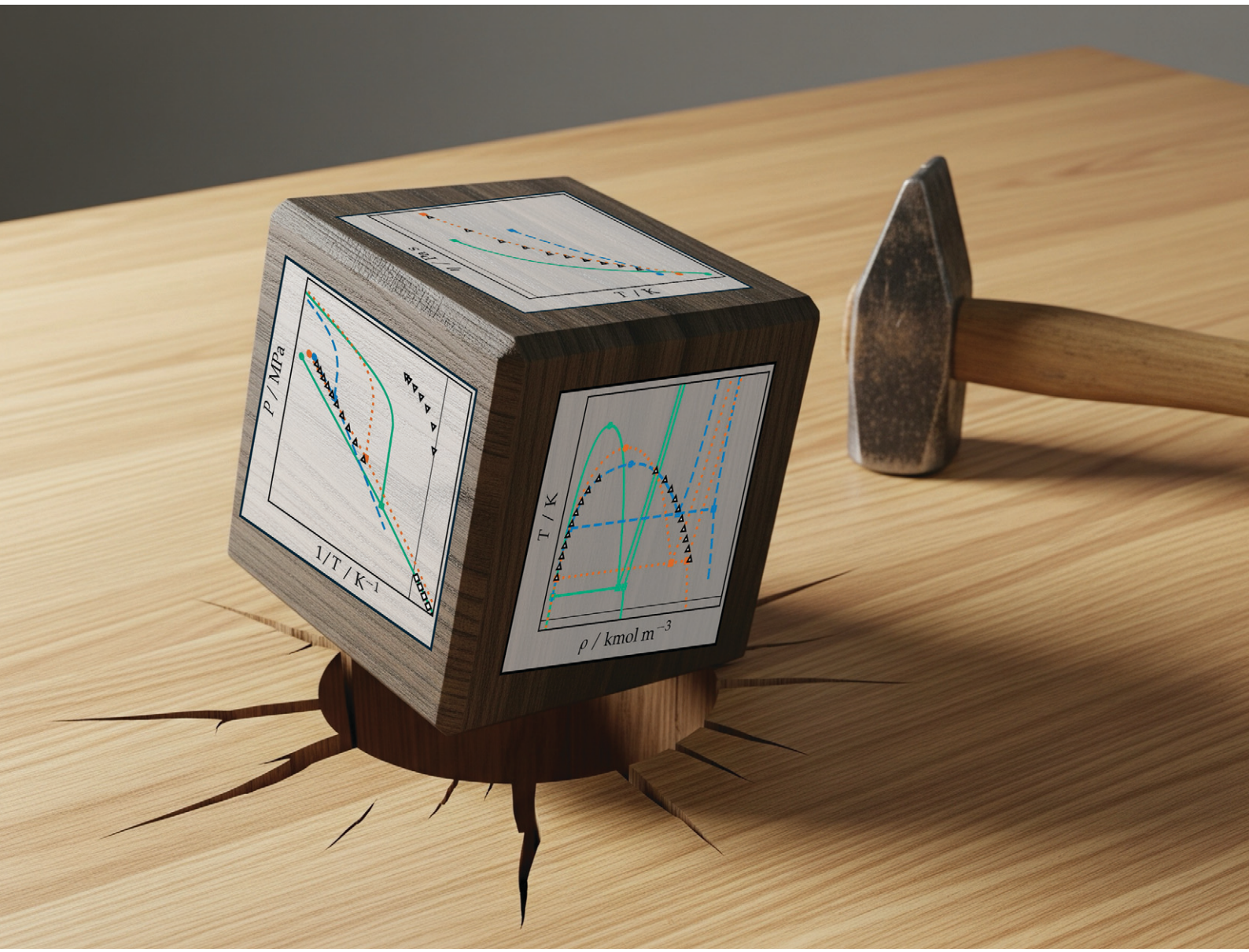
Volume 10
Number 8
August 2025
Pages 601–684

MSDE

Molecular Systems Design & Engineering

rsc.li/molecular-engineering



ISSN 2058-9689

ROYAL SOCIETY OF CHEMISTRY

PAPER

Gustavo Chaparro and Erich A. Müller

Fitting a square peg in a round hole: parameterisation of quasi-spherical molecules employing the Mie potential

IChemE ADVANCING CHEMICAL ENGINEERING WORLDWIDE

MSDE

PAPER

View Article Online
View Journal | View Issue

Cite this: *Mol. Syst. Des. Eng.*, 2025, 10, 620

Fitting a square peg in a round hole: parameterisation of quasi-spherical molecules employing the Mie potential†

Gustavo Chaparro and Erich A. Müller *

The parameterisation of the force field of a molecular system is essential for accurately describing and predicting macroscopic thermophysical properties. Here, we discuss three approaches to obtain the molecular parameters (σ, ε, and λ_r) of the Mie force field from experimental data for quasi-spherical molecules. The first approach is based on a classical strategy that considers fitting only to vapour–liquid equilibria data. The second approach entails a simultaneous fit to equilibrium properties and liquid shear viscosity. Finally, a third approach incorporates solid–fluid equilibrium data. The fitting procedure is facilitated by the use of recently published machine-learned equations of state for the Mie particle, which allows the prediction of thermophysical properties given a set of molecular parameters. The goodness-of-fit is assessed based on the deviations between calculated and experimental data. We also assess the behaviour of the thermal conductivity and speed of sound of the saturated liquid phase to evaluate the transferability of the molecular parameters to properties not used in the parametrisation. Apart from the singular case of monoatomic molecules, no single set of parameters can simultaneously describe the fluid phase equilibria, transport, and solid transition properties of quasi-spherical molecules. This result highlights the limitations of the Mie potential for modelling the thermophysical properties of small molecules. Therefore, a compromise must be made, either to achieve a good description of a specific set of properties or to attain modest accuracy across all phase space.

Received 3rd April 2025,
Accepted 9th June 2025

DOI: 10.1039/d5me00048c

rsc.li/molecular-engineering

Design, System, Application

The fitting of force field parameters to experimental thermophysical data typically relies solely on comparisons with vapour–liquid equilibrium data. We utilise physics-informed machine-learned equations of state to evaluate this classical approach against two alternative parameterisation strategies that incorporate either transport or solid-phase properties. Using the Mie forcefield as a benchmark, it is observed that for simple monoatomic substances, all schemes are essentially equivalent, yielding a good overall fit. However, for more complex molecules, the limitation of the Mie forcefield becomes evident, as no single set of parameters is seen to accurately and simultaneously predict all studied properties. These results highlight the need to develop more robust and detailed force fields for molecular modelling.

1 Introduction

Accurately representing molecular interactions is key for describing and predicting the thermophysical properties of fluids through molecular dynamics (MD) simulations. While, in principle, the intermolecular forces can be resolved *via* the solution of the Schrödinger equations, we are still far from being able to apply these concepts to obtain accurate dispersion energies and comprehensively model fluid phases with accuracy. An accepted approach to circumvent this is to depict the intermolecular interactions *via* mathematical functions known as force fields. Historically, the development of force fields has relied on postulating semi-empirical equations that explicitly model non-bonded, bonded, angular, torsional, and electrostatic interactions.[1–4] These arguably "simple" force fields provide a computationally efficient but incomplete representation of molecular interactions. While successful in many applications, their limitations must be acknowledged,[5] especially when studying systems where polarisation, multi-body effects and detailed electrostatic interactions are significant. Recently, force fields based on machine-learning

Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London, SW7 2AZ, UK.
E-mail: e.muller@imperial.ac.uk

† Electronic supplementary information (ESI) available: Includes details about reduced Mie units and further modelling results of quasi-spherical molecules. See DOI: https://doi.org/10.1039/d5me00048c

(data-driven) techniques[6,7] have been developed to address these challenges. These include force fields based on artificial neural networks,[8,9] Gaussian processes,[10] and many-body expansions.[11,12] Force fields can represent molecules at different resolutions: all-atom (AA), united-atom (UA), and coarse-grained (CG). In AA, all atoms in a molecule are explicitly accounted for, while in UA, small molecular groups are condensed into a single particle; for instance, by treating a $-CH_2-$ group as a single particle. Lastly, in CG, the molecular representation is simplified using larger fictitious particles that group multiple functional groups. CG force fields have been parametrised using experimental thermophysical properties[13] and structural[14,15] data. The CG approach simplifies the molecular representation, allowing the exploration of spatio-temporal scales closer to the macroscopic level.

Regardless of the type of force field (*i.e.*, classical or machine-learning-based) and its molecular resolution (AA, UA, or CG), its applicability depends on the selection of molecular parameters that, along with the state conditions (*i.e.*, density, temperature, and composition), define a Hamiltonian that can be translated into macroscopic observables through molecular simulations and statistical thermodynamics. The appropriate parameterisation of the force field (*i.e.*, the fitting of the molecular parameters) is crucial for its success in predicting thermophysical properties. The parameterisation of a given force field can follow either a bottom-up[16,17] or a top-down[13] approach. In bottom-up approaches, the force field is fitted to reproduce forces, bond distances, and angles from rigorous quantum mechanical calculations. This approach is also the ansatz of machine learning potentials that aim to bridge *ab initio* accuracy with larger scales.[9,18] Conversely, the top-down approach directly optimises the molecular parameters to reproduce relevant experimental data. This method links a molecular model to its implied thermophysical properties and is the most efficient from an engineering and design perspective.

For the sake of conciseness, we shall henceforth discuss the parameterisation of the Mie potential, shown in eqn (1).

$$\begin{aligned} \mathscr{U}^{\mathrm{Mie}} &= \mathscr{C}^{\mathrm{Mie}}\varepsilon\left[\left(\frac{\sigma}{r}\right)^{\lambda_\mathrm{r}} - \left(\frac{\sigma}{r}\right)^{\lambda_\mathrm{a}}\right] \\ \mathscr{C}^{\mathrm{Mie}} &= \frac{\lambda_\mathrm{r}}{\lambda_\mathrm{r}-\lambda_\mathrm{a}}\left(\frac{\lambda_\mathrm{r}}{\lambda_\mathrm{a}}\right)^{\frac{\lambda_\mathrm{a}}{\lambda_\mathrm{r}-\lambda_\mathrm{a}}} \end{aligned} \tag{1}$$

Here, $\mathscr{U}^{\mathrm{Mie}}$ is the interaction energy between two Mie particles, ε is the interaction energy well depth, σ is a characteristic length scale, which is loosely related to the effective particle diameter, and r is the centre-to-centre distance between two monomers. Finally, λ_r and λ_a are the exponents that control the steepness of the repulsive and attractive interactions, respectively. It has been shown that for describing fluid phase equilibria, the exponents of the Mie potential are conformal,[19] implying that multiple combinations of the repulsive and attractive exponents can lead to the same macroscopic equilibrium properties. Following this, for simplicity, the attractive exponent is commonly set to 6 in agreement with the London theory for dispersion forces,[20] resulting in what is sometimes referred to as the (λ_r, 6) Mie potential.[21]

Although the fitting of force fields to experimental data seems a judicious approach, the details of how to proceed remain elusive: which properties should be targeted? Which ones provide the most information? How can one guarantee the robustness and transferability of the optimised parameters? A century ago, Lennard-Jones pioneered the answering of these questions by parameterising the Mie potential using a top-down approach, which led to the well-known Lennard-Jones (LJ) potential (*i.e.*, $U^{\mathrm{LJ}} = 4\varepsilon[(\sigma/r)^{12} - (\sigma/r)^{6}]$).[22,23] He recognised that the molecular parameters of the Mie potential (or the LJ potential) could be fitted to either the second virial coefficient or the viscosity of a dilute argon gas. However, no unique set of parameters could be isolated: "*… so it does not prove possible to obtain a molecular model which will simultaneously explain the two sets of experimental facts*".[23] The results from Lennard-Jones already showed that even for simple molecules like argon, there are trade-offs with respect to which property to choose for fitting.[24]

Within a modern approach, one can target the simultaneous fitting of multiple thermophysical properties and larger data sets. Take, for example, the SAFT-γ Mie force field,[13] a CG force field based on the Mie potential. The parameterisation of this force field is performed by invoking an analytical equation of state (the SAFT-VR-Mie EoS[25]), which accurately maps the force field parameters to macroscopic fluid phase equilibria. By using the equation of state as a surrogate of the underlying MD results, optimised molecular parameters that best fit the vapour–liquid equilibria over a broad range of state points can be effectively determined. The SAFT-γ Mie force field has since been successfully used to model equilibrium properties of alkanes,[26] carbon dioxide,[27] greenhouse gases,[28] water,[29] fluorinated compounds[30,31] and polymers[32–34] amongst others. However, limitations have been seen when attempting to employ the force fields to describe transport properties and to determine the triple point and onset of the solid phases.

In spite of its limitations, the above approach can be generalised by employing the corresponding states principle. Mejía *et al.*[35] developed a corresponding states correlation that relies on the critical temperature (T_c), a saturated liquid density (ρ_l at $T_\mathrm{r} = 0.7$) and the Pitzer acentric factor ($\omega = -1 - \log_{10}(P^{\mathrm{sat}}/P_\mathrm{c})$ at $T_\mathrm{r} = 0.7$) to obtain the molecular parameters of homonuclear Mie chains. This approach, a.k.a. the M&M correlation, has been successfully applied to model the phase equilibria and interfacial properties of a diverse range of industrially relevant molecules, including alkanes, aromatics, gases, and solvents.[36] Moreover, this corresponding states principle has been applied to 6000+ molecular fluids, whose parameters have been freely published in the Bottled SAFT

webpage.[37] Alternatively to the M&M correlation, Hoang *et al.*[38] proposed a similar corresponding states correlation for homonuclear Mie chains. Interestingly, they proposed using the reduced liquid viscosity data point (η_1 at T_r = 0.7) as an additional parameter into the correlation. This approach suggests that transport properties could be employed in addition to the volumetric equilibrium properties to obtain more robust and transferable molecular models.

That being said, this contribution focuses on the single-bead CG modelling of quasi-spherical molecules using the Mie potential, eqn (1). Three main parameterisation approaches are explored and compared. The first approach follows the classical strategy for parameterising force fields where only vapour–liquid equilibria and fluid data are considered. The second approach follows the corresponding states principle of Hoang *et al.*,[38] where the liquid shear viscosity data is considered in addition to the VLE data. The third parameterisation approach includes solid–fluid data along with fluid phase equilibria. Details of the different parameterisation approaches and how the thermophysical properties are modelled are discussed in section 2, while the discussion of the results is given in section 3. Finally, the conclusions are summarised in section 4.

2 Multi-property parametrisation approaches

Quasi-spherical molecules can be described using the Mie potential, eqn (1). Upon setting the attractive exponent to 6, the Mie isotropic particle is fully specified by three molecular parameters (σ, ε, and λ_r). These molecular parameters must be fitted to experimental data for the potential to represent thermophysical properties.

Here, we consider three parameterisation approaches to obtain the molecular parameters (σ, ε, λ_r) of the Mie potential. We formulate a general objective function as follows.

$$\mathrm{OF}(\sigma,\ \varepsilon,\ \lambda_r) = \sum_k \frac{w_k}{N_{p,k}} \sum_{i=1}^{N_{p,k}} \left| \frac{y_{k,i}^{\mathrm{pred}}}{y_{k,i}^{\mathrm{exp}}} - 1 \right| \quad (2)$$

This objective function, eqn (2), represents a weighted average of relative deviations of various thermophysical properties "k". In this equation, w_k is the relative weight, $N_{p,k}$ is the number of data points, and y_k^{pred} and y_k^{exp} are the predicted and experimental values of the property "k". A benchmark objective function (OF_1), and arguably the most common strategy for parameterising the Mie potential, is to exclusively consider properties related to vapour–liquid equilibria (VLE), such as saturation pressure (P_{VLE}), saturated liquid density ($\rho_{\mathrm{l,VLE}}$), and vaporisation enthalpy (ΔH_{VLE}). This first objective function resembles the parametrisation of SAFT EoSs.[25,39] The second objective function (OF_2) includes the liquid shear viscosity (η) alongside the VLE data. Finally, the third objective function (OF_3) considers the properties of the VLE while also incorporating the data of the solid–fluid equilibria. The latter objective function allows one to assess the impact of the solid phase on the molecular parameters. The solid phase data includes the melting pressure (P_{SLE}), sublimation pressure (P_{SVE}) and sublimation enthalpy (ΔH_{SVE}). The details of the objective functions are summarised in Table 1. The relative weight (w_k) for all thermophysical properties in the objective function is set at 1, except for the melting pressure, whose relative weight is set to w_k = 0.01. This relatively small weight is chosen to prevent biasing the objective function, as melting pressures are orders of magnitude larger than those of all other considered values.

Table 1 Parameterisation approaches to obtain the molecular parameters of the Mie potential[a]

	P_{VLE}[b]	$\rho_{\mathrm{l,VLE}}$[b]	ΔH_{VLE}[b]	P_{SLE}[c]	P_{SVE}[d]	ΔH_{SVE}[d]	η[e]
OF_1	✓	✓	✓	✗	✗	✗	✗
OF_2	✓	✓	✓	✗	✗	✗	✓
OF_3	✓	✓	✓	✓	✓	✓	✗

[a] Thermophysical properties considered in the objective function (OF) are indicated by a "✓". [b] Vapour–liquid equilibria (VLE) properties are also referred to as "vaporisation" properties. These include the vaporisation pressure (P_{VLE}), saturated liquid density ($\rho_{\mathrm{l,VLE}}$), and vaporisation enthalpy (ΔH_{VLE}). VLE properties can be obtained either with the FE-ANN EoS[40] or FE-ANN(s) EoS,[41] based on eqn (3). [c] Solid–liquid equilibria (SLE) properties are also referred to as "melting" properties. Only the melting pressure (P_{SLE}) is considered here. [d] Solid–vapour equilibria (SVE) properties are also referred to as "sublimation" properties. These include the sublimation pressure (P_{SVE}) and sublimation enthalpy (ΔH_{SVE}). SLE and SVE properties are obtained using the FE-ANN(s) EoS. [e] The shear viscosity (η) is modelled using the ANN-based models developed in ref. 42, based on eqn (6).

2.1 Experimental data acquisition

The objective functions based on eqn (2) require experimental data. Pseudo-experimental data is obtained from the NIST TRC database.[43] The data reported by NIST have already been curated and fitted to accurate empirical functions. The use of the data sets provided by NIST provides advantages over the use of discrete experimental data sets. Not only is the pseudo-data smooth and devoid of outliers, but it is also "continuous" in nature and allows for evaluating properties over an extensive range of state points. Furthermore, the data is internally and thermodynamically consistent.

Thermophysical properties under VLE conditions are generated within the range $T_r = T/T_c \in [0.55, 0.95]$. The VLE data includes saturation pressure, saturated liquid density, vaporisation enthalpy, and saturated liquid viscosity. In addition, the thermal conductivity and speed of sound of the saturated liquid phase are generated to assess the transferability of the molecular parameters to unseen thermophysical properties. Similarly, the melting pressure data is generated in the $T_r = T/T_c \in [0.55, 1.1]$ range, and the

sublimation pressure and enthalpies are generated in the $T_r = T/T_c \in [0.45, 0.55]$ range. These constitute our database of experimental information.

2.2 Thermophysical properties modelling

To assess the objective functions, we are additionally required to evaluate the thermophysical properties corresponding to the Mie potential at a given state point as a function of the molecular parameters (σ, ε, and λ_r). The thermophysical properties of the Mie particle are modelled using physics-informed Free Energy Artificial Neural Networks (FE-ANNs). We employ the FE-ANN EoS[40] that explicitly models the Helmholtz free energy of the Mie fluid and can predict the fluid phase properties and VLE with an accuracy that matches MD results. Similarly, we employ the FE-ANN(s) EoS[41] that extends the framework to handle solid–fluid equilibria, such as SLE and SVE. The Helmholtz free energy is a convenient state function for computing thermophysical properties and phase equilibria, and the reader is referred to ref. 40, 41 and 44 for further details about Helmholtz free energy thermodynamics. The FE-ANN/FE-ANN(s) EoSs are formulated using reduced units, indicated by a "*" superscript. The framework can be used for actual molecules by converting the results into real units, as shown below. The reader is referred to section S.1 of the ESI† for further details about reduced units.

$$\begin{aligned} A^{res} &= N_{Av}\varepsilon A^{*,res} \\ A^{*,res} &= \mathrm{ANN}\left(\alpha_{vdw}, \rho N_{Av}\sigma^3, \frac{\varepsilon}{Tk_B}\right) \\ &\quad - \mathrm{ANN}\left(\alpha_{vdw}, \rho = 0, \frac{\varepsilon}{Tk_B}\right) \end{aligned} \tag{3}$$

In eqn (3), A^{res} is the residual Helmholtz free energy (in J mol^{-1}), ρ is the molar density (in mol m^{-3}), T is the temperature (in K). The shape parameter (σ, in m) and energy scale (ε, in J) are the molecular parameters. Finally, N_{Av} and k_B are the Avogadro and Boltzmann constants.[45] The Mie potential is also described through the α_{vdw} parameter. This parameter is defined below for (λ_r, $\lambda_a = 6$) Mie particles.

$$\alpha_{vdw}(\lambda_r, \lambda_a = 6) = \mathscr{C}^{Mie}\left[\frac{1}{3} - \frac{1}{\lambda_r - 3}\right] \tag{4}$$

where, $\mathscr{C}^{Mie}$ is defined in eqn (1). Then, the total Helmholtz free energy is obtained by adding residual Helmholtz free energy obtained from the FE-ANN EoS and the ideal gas contribution obtained from statistical thermodynamics.[24]

$$A = A^{res} + \underbrace{RT\left[\ln\left(\Lambda^3\rho\right) - 1\right]}_{A^{id}} \tag{5}$$

Here, the ideal contribution (A^{id}) depends on the ideal gas constant (R), temperature (T), density (ρ) and the de Broglie volume (Λ^3, where $\Lambda \propto 1/\sqrt{T}$).

The second objective function (OF_2) requires predicting the shear viscosity of the Mie fluid. For this purpose, we employ the artificial neural network (ANN)-based models for the transport properties of the Mie fluid reported in ref. 42. The recommended model for the shear viscosity (η) is based on an ANN and utilises a semi-log scale. This model is formulated using reduced units and can be converted into real units, as shown below.

$$\begin{aligned} \eta &= \frac{\sqrt{\varepsilon M/N_{Av}}}{\sigma^2}\exp[\ln\eta^*] \\ \ln\eta^* &= \mathrm{ANN}\left(\alpha_{vdw}, \rho N_{Av}\sigma^3, \frac{\varepsilon}{Tk_B}\right) \end{aligned} \tag{6}$$

Here, η is the shear viscosity (in Pa s) and M is the molar mass (in kg mol^{-1}).

The behaviour of the speed of sound and the thermal conductivity will also be studied to assess the transferability of the optimised molecular parameters (σ, ε, λ_r) to other thermophysical properties. The speed of sound can be obtained by any EoS as follows.

$$w = \sqrt{\left[\rho\kappa_T\left(\frac{C_V}{C_P}\right)M\right]^{-1}} \tag{7}$$

Here, w is the speed of sound (in m s^{-1}), κ_T is the isothermal compressibility (in Pa^{-1}) and C_V and C_P are obtained isochoric and the isobaric heat capacities (in J K^{-1} mol^{-1}). The FE-ANN/FE-ANN(s) EoSs explicitly model the residual Helmholtz free energy and the ideal contribution can be accounted for analytically as shown in eqn (5). The isothermal compressibility is directly obtained from the Helmholtz free energy as $\kappa_T = (\partial\rho/\partial P)_T/\rho$. It is recommended that the ideal and residual contributions for the heat capacities be considered separately.

$$C_V = C_V^{id} + C_V^{res} \tag{8a}$$

$$C_P = C_P^{id} + C_P^{res} \tag{8b}$$

Here, the residual contributions, indicated by a "res" superscript, are obtained from the derivatives of the FE-ANN/FE-ANN(s) EoSs. The ideal contributions (C_V^{id} and C_P^{id}) are computed from the correlations from the DIPPR project 801.[46] These correlations consider rotational and vibrational contributions to the ideal gas heat capacities.

Finally, the thermal conductivity is obtained from the recommended ANN-based model developed in ref. 42. This model uses reduced units and a semi-log scale. Therefore, the thermal conductivity of the Mie fluid is obtained as follows.

$$\begin{aligned} \kappa^{ANN} &= \frac{k_B\sqrt{N_{Av}\varepsilon/M}}{\sigma^2}\exp[\ln\kappa^*] \\ \ln\kappa^* &= \mathrm{ANN}\left(\alpha_{vdw}, \rho N_{Av}\sigma^3, \frac{\varepsilon}{Tk_B}\right) \end{aligned} \tag{9}$$

In eqn (9), κ^{ANN} is the thermal conductivity (in W m^{-1} K^{-1}). This model is valid for a wide range of density and

temperature conditions. However, it predicts an incorrect trend at low temperatures (close to the triple point) across the saturation VLE line.[41] Further details about this numerical artefact are shown in Fig. S.1 of the ESI.† This ANN-based model considers the thermal conductivity from non-bonded interactions. Real molecules also exhibit a thermal conductivity contribution due to the bond vibrations. The ANN-based model does not consider these contributions. The thermal conductivity of polyatomic atomic molecules can be obtained using Eucken formalism.[47] However, this approach is limited to dilute states.[48] Instead, we account for the vibrational contribution using the correction term proposed by Liang and Tsai.[49]

$$\kappa^{\text{vib}} = \rho D\left(C_{\text{V}}^{\text{id}} - \#\text{ non-vibrational degrees of freedom}\cdot\frac{R}{2}\right) \quad (10)$$

The correction term considers the self-diffusivity (as ρD), which can be obtained from the ANN-based model in ref. 42, as shown below.

$$\begin{aligned}\rho D &= \frac{1}{\sigma^2}\sqrt{\frac{\varepsilon}{MN_{\text{Av}}}}\rho^* D^* \\ \rho^* D^* &= \text{ANN}\left(\alpha_{\text{vdw}}, \rho N_{\text{Av}}\sigma^3, \frac{\varepsilon}{Tk_{\text{B}}}\right)\end{aligned} \quad (11)$$

The ideal gas isochoric heat capacity is obtained from the correlations of the DIPPR project 801.[46] The number of non-vibrational degrees of freedom of a molecule depends on its molecular structure. Monoatomic molecules have three non-vibrational degrees of freedom. Linear molecules, such as carbon monoxide, have five non-vibrational degrees of freedom. Non-linear molecules, like methane, have six non-vibrational degrees of freedom. The correction term κ^{vib} is small for liquid phases. For the molecules of interest of this work $\kappa^{\text{vib}}/\kappa^{\text{ANN}} \lesssim 10^{-2}$. Finally, the thermal conductivity is obtained by adding eqn (9) and (10).

$$\kappa = \kappa^{\text{ANN}} + \kappa^{\text{vib}} \quad (12)$$

2.3 Multi-start optimisation approach

The objective functions, based on eqn (2), are minimised to find suitable parameters (σ, ε, λ_{r}) to describe the phase behaviour of a given molecule. The objective function is minimised using the Nelder–Mead method[50] included in the minimize function from the scipy.optimize Python library.[51] The Nelder–Mead method, also known as the simplex method, is a derivative-free local minimisation algorithm capable of handling bounded optimisation. The optimised molecular parameters are highly dependent on the initial guess as the Nelder–Mead method can easily get stuck on a local solution of the optimisation surface. The optimisation could be initialised from molecular parameters obtained from the corresponding states principle correlations.[35,38] However, these initial guesses can potentially bias the optimisation results. In order to find an optimal "global" set of molecular parameters, a multi-start optimisation approach using 32 initial guesses generated through Sobol' sequences[52] is used here. This number of initial guesses was chosen to enhance the likelihood of reaching multiple local optimal solutions, from which the one with the lowest objective value is selected. The ranges of the initial guesses are detailed in Table 2. Furthermore, to operate in the validity region of the ANN-based models, the repulsive exponent is restricted to be in the $\lambda_{\text{r}} \in [8, 34]$ range.

Table 2 Optimisation details to fit the Mie potential molecular parameters

		Initial guess[a]		Optimisation bounds[b]	
	Units	Min value	Max value	Min value	Max value
σ	Å	2.5	3.5	—	—
ε/k_{B}	K	100	300	—	—
λ_{r}	—	9	26	8	34

[a] 32 initial guesses are generated in these ranges using Sobol' sequences.[52] [b] The repulsive exponent is bounded to be in the validity region of the ANN-based models.

3 Parameterisation results for quasi-spherical molecules

The molecular parameters of the Mie potential have been determined for representative quasi-spherical molecules. The selected molecules are sufficient to assess the limitations of the Mie potential and to provide insights into good practices for its parameterisation. The selected quasi-spherical molecules can be categorised into three major groups: monoatomic (argon, krypton, and xenon), linear (nitrogen and carbon monoxide), and tetrahedral (methane and tetrafluoromethane) molecules.

The optimised molecular parameters and their corresponding deviations are summarised in Table 3. This table also includes the parameters predicted by the M&M[35] and the Hoang *et al.*[38] corresponding states correlations. Overall, it is evident from this table that the VLE can be modelled with low deviations (MAPE $\lesssim$ 5%) when it is the sole target data (*i.e.*, OF_1). It must be noted that the parameters obtained from this objective function are in good agreement with the M&M correlation. This similarity results from the M&M correlation using the critical temperature, acentric factor, and liquid density to predict molecular parameters that correctly predict the VLE conditions at $T_{\text{r}} = 0.7$. However, this correlation is parametrised to predict a correct density behaviour, and higher deviations (>10%) are observed for the saturation pressure (P_{VLE}). This higher deviation in pressure is also observed from the parameters obtained in the correlation of Hoang *et al.*[38] This correlation targets a correct density and shear viscosity behaviour. The objective of the parameters provided by this correlation is comparable to the ones obtained from using OF_2.

Table 3 Optimised Mie molecular parameters for selected quasi-spherical molecules[a,b]

					MAPE[c,d]								
Component	[i]	σ[e]	ε/k_B[e]	λ_r[e]	P_{VLE}	$\rho_{l,VLE}$	ΔH_{VLE}	P_{SLE}	P_{SVE}[f]	ΔH_{SVE}[f]	η	κ[g,h]	w[h]
Argon	M&M[35]	3.41	126.70	14.34	11.2	1.65	1.58	9.98	6.10	1.67	2.55	2.86	**1.43**
	Hoang *et al.*[38]	3.41	125.55	13.92	7.53	**1.33**	**0.57**	5.91	4.61	1.33	1.38	2.17	2.52
	OF_1	3.43	124.98	13.52	1.13	2.71	3.29	5.75	4.43	1.57	0.85	**1.16**	4.59
	OF_2	3.44	123.19	13.16	**0.90**	3.31	2.70	4.21	**0.34**	**0.57**	1.80	1.24	4.56
	OF_3	3.42	124.62	13.53	1.90	2.04	2.23	**3.68**	0.35	1.20	**0.57**	1.29	3.95
Krypton	M&M[35]	3.63	172.12	13.72	13.4	1.58	1.71	**5.13**	16.7	0.46	1.63	15.5	**1.87**
	Hoang *et al.*[38]	3.64	171.81	13.52	9.49	**1.28**	**0.27**	5.24	12.1	0.64	**1.38**	15.0	2.92
	OF_1	3.67	172.85	13.37	0.92	2.82	3.42	8.47	2.92	1.69	2.52	14.4	5.12
	OF_2	3.68	169.42	12.89	**0.59**	3.86	2.56	8.69	3.79	**0.31**	1.85	16.3	4.98
	OF_3	3.66	172.84	13.44	1.77	2.13	2.63	7.65	**0.36**	1.53	2.39	**14.3**	4.58
Xenon	M&M[35]	3.96	246.86	14.73	13.8	1.62	1.85	19.7	10.0	5.89	3.31	5.67	**3.70**
	Hoang *et al.*[38]	3.96	244.38	14.25	9.87	**1.30**	**0.60**	13.5	9.23	5.37	**1.94**	5.08	4.55
	OF_1	4.00	245.07	14.00	0.93	3.21	3.88	17.4	19.0	6.34	2.62	4.48	7.32
	OF_2	4.01	238.49	13.28	**0.42**	4.02	2.98	**4.46**	**7.13**	4.10	2.18	2.23	5.96
	OF_3	4.00	235.37	12.96	1.00	3.39	2.48	7.05	7.34	**3.39**	3.18	**1.76**	5.83
Nitrogen	M&M[35]	3.65	120.65	19.11	11.9	1.59	2.82	69.2	44.8	10.9	17.3	**7.81**	4.47
	Hoang *et al.*[38]	3.61	106.10	14.16	15.8	**1.24**	4.76	52.2	8.89	0.75	**1.49**	16.6	**3.66**
	OF_1	3.66	115.19	16.30	**1.15**	1.89	**2.67**	63.7	39.0	7.79	15.5	9.57	6.03
	OF_2	3.67	105.83	13.57	3.03	4.76	2.69	53.8	11.1	1.58	2.70	17.0	4.56
	OF_3	3.62	101.64	12.39	4.07	1.71	3.95	**46.5**	**1.44**	**0.59**	5.31	16.9	6.70
Carbon monoxide	M&M[35]	3.68	130.95	20.67	11.8	1.16	2.31	116.6	45.5	8.07	15.7	3.51	7.22
	Hoang *et al.*[38]	3.66	123.23	17.39	10.0	**0.86**	**2.01**	120.4	24.9	3.17	15.2	6.39	**3.92**
	OF_1	3.69	126.29	17.89	**0.98**	1.46	2.65	119.2	42.3	5.77	13.2	**3.07**	8.70
	OF_2	3.63	115.70	14.20	3.60	2.34	3.72	129.1	16.7	**0.31**	**12.3**	7.09	5.61
	OF_3	3.65	108.96	12.60	5.38	1.91	4.45	**83.7**	**3.21**	3.88	17.4	13.2	5.88
Methane	M&M[35]	3.74	162.71	14.79	14.3	2.16	3.65	65.8	27.0	9.12	19.6	5.48	**1.40**
	Hoang *et al.*[38]	3.71	138.42	11.06	27.9	**1.67**	7.30	**30.8**	54.2	1.93	**1.83**	11.3	8.99
	OF_1	3.78	162.32	14.20	**0.50**	3.35	**3.50**	66.1	38.8	9.93	24.8	**3.16**	5.93
	OF_2	3.82	139.81	10.83	5.25	7.48	4.58	38.3	9.82	**0.64**	2.47	12.3	9.11
	OF_3	3.76	139.96	10.68	4.62	3.29	5.55	35.9	**3.96**	0.89	6.92	8.31	10.0
Tetrafluoro-methane	M&M[35]	4.38	269.16	38.29	16.0	5.98	11.2	2858.6	—	—	31.3	14.1	**7.42**
	Hoang *et al.*[38]	4.17	184.24	13.23	69.4	**1.33**	22.0	915.0	567.8	13.7	**3.00**	33.3	18.6
	OF_1	4.37	266.77	34.00	**4.10**	1.97	**1.55**	4415.4	—	—	44.8	**9.52**	10.1
	OF_2	4.40	191.17	12.59	17.0	11.0	12.4	901.0	114.9	**6.93**	7.00	32.7	14.5
	OF_3	5.39	153.76	8.00	68.2	49.4	15.3	**341.7**	**14.8**	15.6	36.9	55.1	21.4

[a] Thermophysical properties included in the objective functions are coloured with a light grey background. Refer to eqn (2) and Table 1 for further details about OF_1, OF_2, and OF_3. [b] The objective functions are optimised using pseudo-experimental data obtained from the NIST TRC.[43] [c] The reported Mean Absolute Percentage Error (MAPE) is obtained as $\mathrm{MAPE} = 100/N_P \sum_{i=1}^{N_P} \left| y_i^{\mathrm{model}}/y_i^{\mathrm{true}} - 1 \right|$. The reported MAPE include the vaporisation pressure (P_{VLE}), saturated liquid density ($\rho_{l,VLE}$), vaporisation enthalpy (ΔH_{VLE}), melting pressure (P_{SLE}), sublimation pressure (P_{SVE}), sublimation enthalpy (ΔH_{SVE}), shear viscosity (η), thermal conductivity (κ), and speed of sound (w). [d] The smallest MAPE deviations per component are highlighted in **bold**. [e] σ, ε/k_B, and λ_r are the molecular parameters of the Mie potential, eqn (1). λ_a is set to 6. σ is reported in Angstroms (1 Å = 10^{-10} meters) and ε/k_B is reported in Kelvin. [f] No MAPE is reported for the sublimation pressure and enthalpy when the pseudo-experimental data is below the validity of the FE-ANN(s) EoS, *i.e.*, $T^* < 0.6$. [g] The predicted thermal conductivity is obtained as in eqn (12) which is based on eqn (9) and (10). [h] The modelled thermal conductivity and speed of sound use the ideal gas heat capacities correlations from the DIPPR project 801.[46] [i] Molecular parameters obtained from the M&M[35] and Hoang *et al.*[38] correlations are based on the corresponding states principle. The M&M correlation relies on the critical temperature, acentric factor, and liquid density at $T_r = 0.7$. The correlation of Hoang *et al.*[38] additionally uses the saturated liquid viscosity at $T_r = 0.7$.

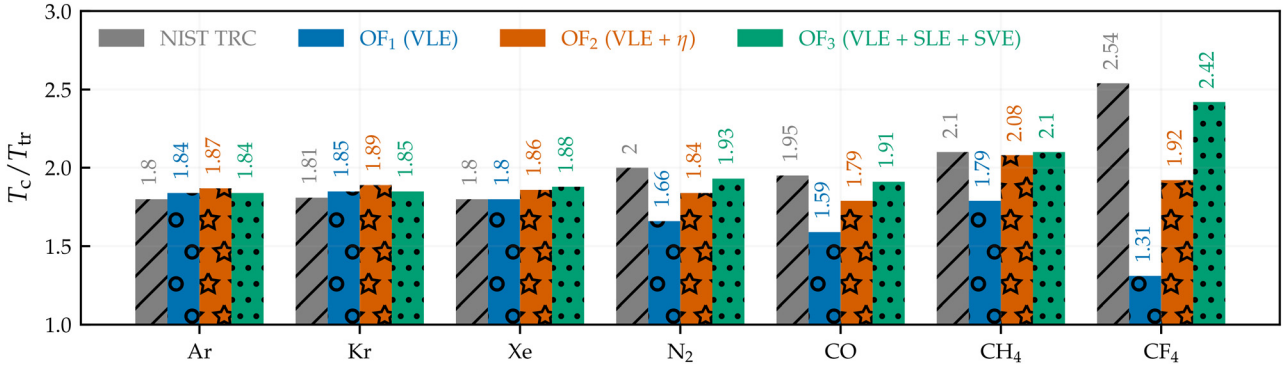


Fig. 1 Critical and triple temperature ratio of selected quasi-spherical molecules. Grey bars with diagonal hatch refer to the value obtained from the NIST TRC database.[43] Blue bars with a circle pattern refer to the values obtained from OF_1 (VLE only). Orange bars with a star pattern refer to the values obtained from OF_2 (VLE + shear viscosity). Green with a dotted pattern refers to the values obtained from OF_3 (VLE + SLE + SVE).

Using OF_1 (or the M&M correlation) provides a good description of the VLE. However, the overfitting of the VLE phase envelope commonly comes at the cost of increasing the repulsions and interaction energy of the Mie potential. As will be shown later in Fig. 3–5, the apparent excellent agreement with the experiments stems from extending the validity range of a fluid EoS to regions below the triple temperature where a solid phase would exist. This overfitting, in itself, is not problematic if only fluid properties are considered, but it has implications if one wishes to transfer the model to other thermophysical properties. For example, an erroneous prediction of triple points will substantially influence viscosity predictions, which are known to rise exponentially as one approaches solid-phase conditions.[53]

The ratio between the critical (T_c) and triple (T_{tr}) points for the selected molecules is illustrated in Fig. 1. From this figure, it is apparent that T_c/T_{tr} is accurately represented for monoatomics when using only VLE data. However, including data on shear viscosity (OF_2) or solid–fluid equilibria (OF_3) data is essential to more effectively predict this ratio for polyatomic molecules. The proposed parameterisation approaches can also predict other transport properties, such as the self-diffusivity coefficient. As shown in Fig. S.5 in the ESI,† the self-diffusivity can be accurately predicted by the three different objective functions as long as the stable fluid region is correctly predicted (*i.e.*, correct T_c/T_{tr} ratio). This behaviour results from the dependence of the self-diffusivity on the molecular size (related to σ) for non-associating molecules. This result has been observed in the early[13] and recent[54] parameterisation of the SAFT-γ force field and implies that this property is not determinant when parameterising coarse-grained force fields that ignore explicit polar moments and electrostatics. These general observations set the stage for the main discussion below, which explains how achieving acceptable fits for certain properties comes at the price of producing unreliable representations of others. The discussion will be presented below based on the aforementioned molecular groups (*i.e.*, monoatomic, linear, and tetrahedral molecules). Further results for molecules not covered in this section are available in section S.2 of the ESI.†

3.1 Monoatomic molecules

Three monoatomic molecules are considered here: argon (Ar), krypton (Kr), and xenon (Xe). In Table 3, we show how the fitted characteristic parameters of each molecule are similar regardless of which objective function is used. Consequently, the deviations of each thermophysical property fall within the same ranges. Nevertheless, interesting insights emerge when considering transport properties or solid–fluid equilibria data. It can be seen that including the shear viscosity in the objective function (OF_2) tends to reduce the deviations in the properties related to SLE and SVE. It is understood that the shear viscosity and the description of the solid phase are sensitive to short distances between molecules where the repulsive part of an interaction potential becomes relevant. Thus, improving the description of one property benefits the other. It can further be noted that the molecular parameters obtained can correctly predict the thermal conductivity and speed of sound of these molecules. The thermal conductivity is accurately predicted for argon and xenon (MAPE $\kappa \lesssim 4\%$). The thermal conductivity for krypton, however, is predicted with higher deviations, on the order of 15%. Similarly, the Mie potential can predict the speed of sound with relative errors in the range of 5% for these spherical molecules. The speed of sound serves as a stringent test for a set of molecular parameters, as it depends on multiple derivatives of the Helmholtz free energy. Moreover, the speed of sound can be accurately measured,[55,56] enabling an assessment of the limitations of the molecular model and its parameters.

The results of Table 3 are best interpreted graphically. The phase equilibria and selected thermophysical properties results of argon are illustrated in Fig. 2. Similar figures for krypton and xenon can be found in the ESI.† Fig. 2 takes into account the three sets of parameters along with two possible modelling approaches. Fig. 2(a)–(c) depict the phase equilibria results modelled solely with fluid-phase data. Generally, good agreement can be noted with the reference VLE data when using the OF_1 and

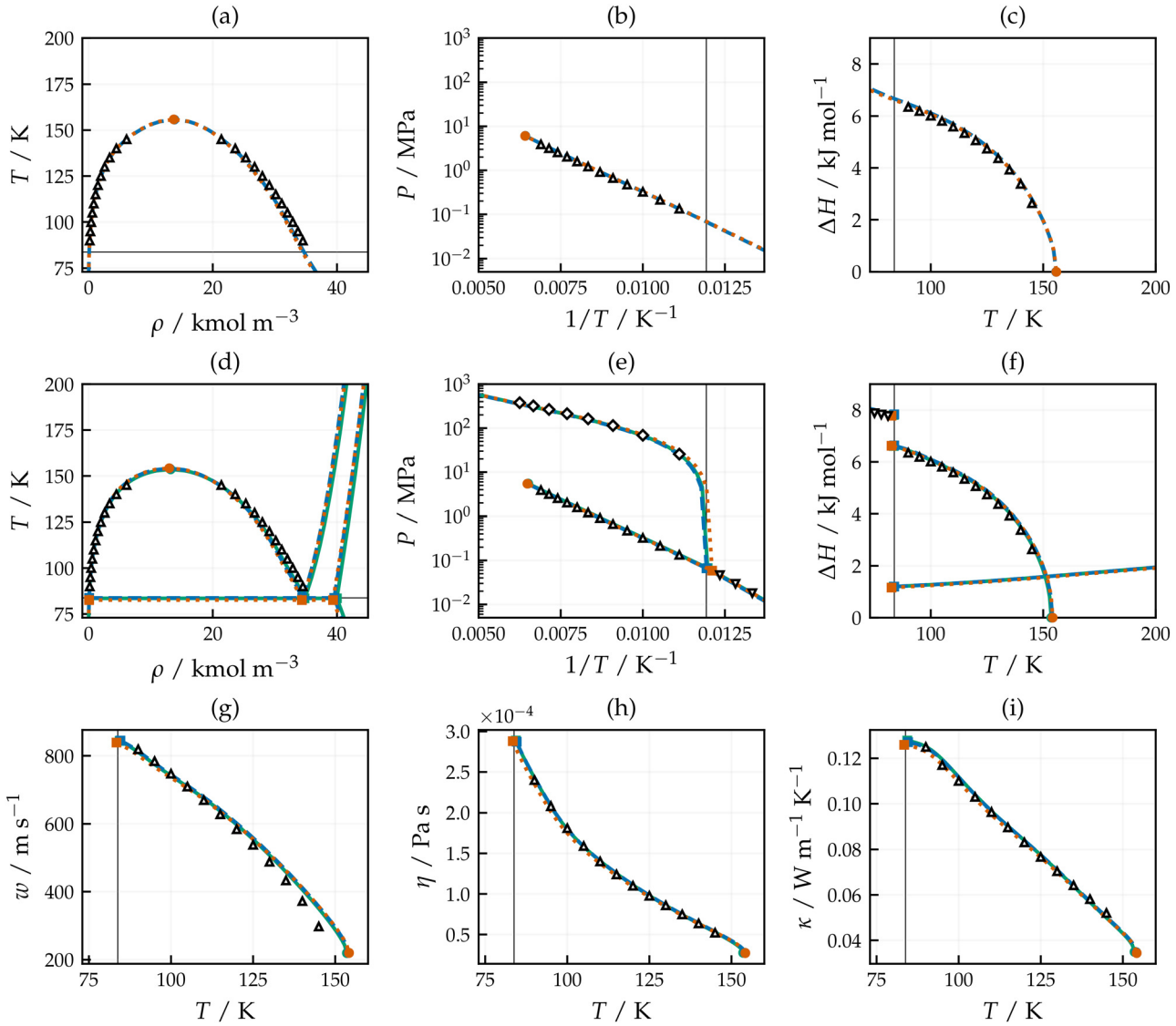


Fig. 2 Selected thermophysical properties argon. (a) and (d) Phase envelope. (b) and (e) Clapeyron plot. (c) and (f) Enthalpy of phase change. (g) Speed of sound. (h) Shear viscosity. (i) Thermal conductivity. (a)–(c) Phase equilibria modelled using the FE-ANN EoS developed ref. 40. (d)–(g) Phase equilibria and speed of sound modelled using the FE-ANN(s) EoS developed in ref. 41. FE-ANN/FE-ANN(s) EoSs are based on eqn (3). (h) and (i) The shear viscosity and thermal conductivity are modelled using the ANN-based models developed ref. 42, based on eqn (6) and (12). Dashed blue lines and symbols refer to results using parameters optimised with OF_1 (VLE only). Dotted orange lines and symbols refer to results using parameters optimised with OF_2 (VLE + shear viscosity). Solid green lines and symbols refer to results using parameters optimised with OF_3 (VLE + SLE + SVE). Refer to eqn (2) and Table 1 for further details about the objective functions. Filled circle is the critical point. Filled square is the triple point. Pseudo-experimental data obtained from NIST TRC.[43] Upright triangles are VLE data, diamonds are SLE data, and downward triangles are SLE data. For reference, the solid black line refers to the triple temperature.

OF_2 parameterisation approaches. Fig. 2(d)–(f) examine the same set of molecular parameters but using the FE-ANN(s) EoS. This EoS broadens the prediction to encompass the entire phase diagram, including VLE, SLE, SVE, and the triple point. It can be observed that even when solid-phase data are excluded from the parameterisation, all modelling approaches converge to predict a similar phase envelope, equilibrium pressure, and enthalpy of phase change. These sets of parameters also accurately predict the triple temperature, which is marked by a solid black line in the figures.

The results for the speed of sound, shear viscosity and thermal conductivity of saturated liquid argon are presented in Fig. 2(g)–(i). Similarly to the phase equilibria results, the three parameterisation approaches accurately predict these thermophysical properties. These findings further confirm that the Mie potential can model the thermophysical properties of simple monoatomics with precision. Notably, the optimal repulsive exponents for these molecules are closer to the value of 13–13.5 than to the "classical" value of 12 employed in the LJ model.

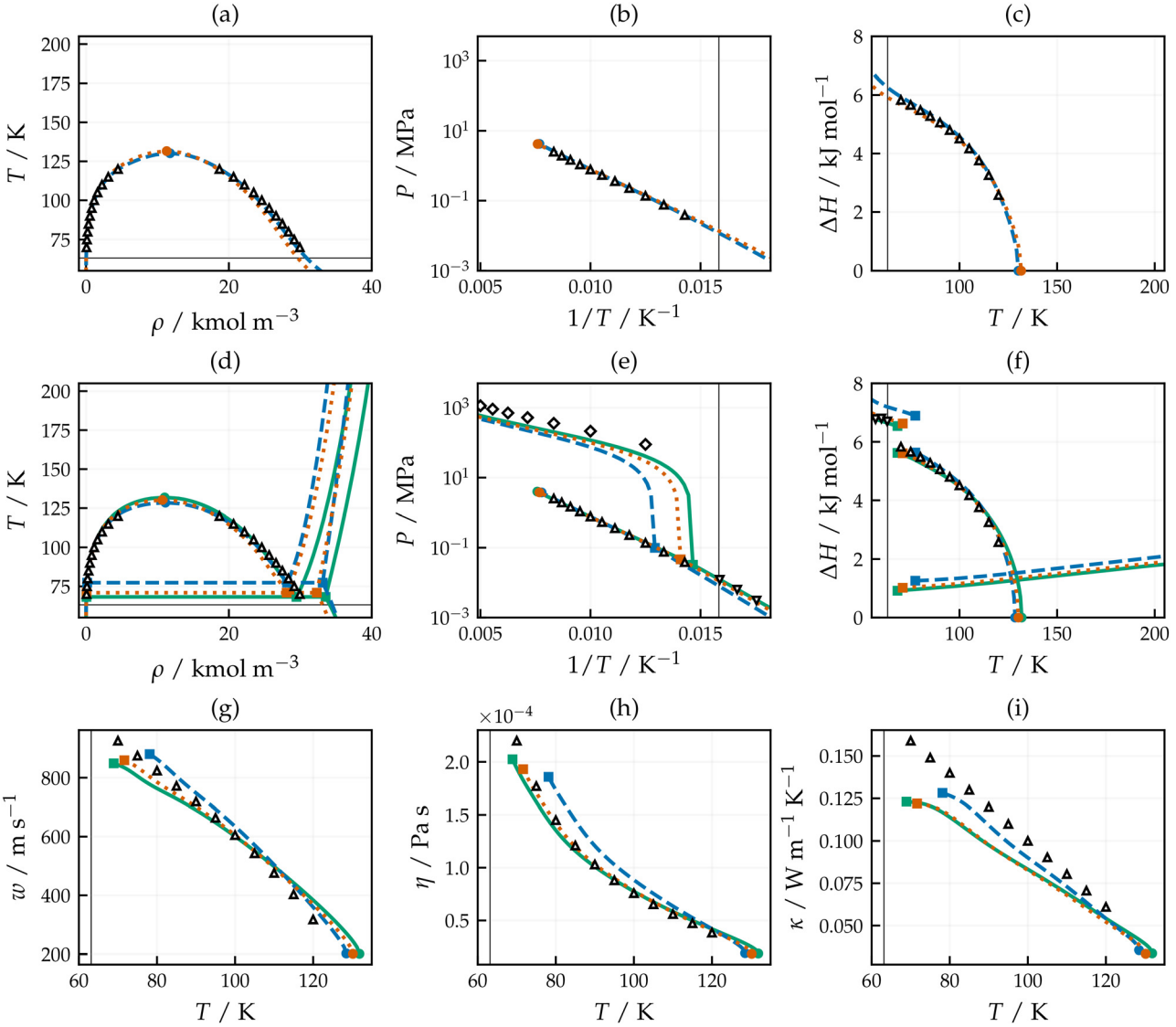


Fig. 3 Phase equilibria modelling of nitrogen. (a) and (d) Phase envelope. (b) and (e) Clapeyron plot. (c) and (f) Enthalpy of phase change. (g) Speed of sound. (h) Shear viscosity. (i) Thermal conductivity. (a)–(c) Phase equilibria modelled using the FE-ANN EoS developed ref. 40. (d)–(g) Phase equilibria and speed of sound modelled using the FE-ANN(s) EoS developed in ref. 41. FE-ANN/FE-ANN(s) EoSs are based on eqn (3). (h) and (i) The shear viscosity and thermal conductivity are modelled using the ANN-based models developed ref. 42, based on eqn (6) and (12). Dashed blue lines and symbols refer to results using parameters optimised with OF_1 (VLE only). Dotted orange lines and symbols refer to results using parameters optimised with OF_2 (VLE + shear viscosity). Solid green lines and symbols refer to results using parameters optimised with OF_3 (VLE + SLE + SVE). Refer to eqn (2) and Table 1 for further details about the objective functions. Filled circle is the critical point. Filled square is the triple point. Pseudo-experimental data obtained from NIST TRC.[43] Upright triangles are VLE data, diamonds are SLE data, and downward triangles are SLE data. For reference, the solid black line refers to the triple temperature.

3.2 Linear molecules

While the various parameterisation choices did not play a significant role in determining the molecular parameters of noble gases, a different scenario unfolds as the underlying molecules become more complex. This section will discuss the effects of the parameterisation approach on small linear molecules, such as nitrogen (N_2) and carbon monoxide (CO). From the results shown in Table 3 for these molecules, it can be observed that the VLE properties are better predicted with OF_1 albeit at the cost of increased repulsiveness ($\uparrow\lambda_r$) and energy ($\uparrow\varepsilon$) of the interactions. The size parameter, σ, which is closely related to the liquid density, remains consistent across the three parameterisation approaches. Including the shear viscosity (OF_2) or solid (OF_3) data in the optimisation results in a decrease in the repulsive exponent and interaction energy of the potential with these two properties targeting molecular parameters in the same direction. It can be observed that there is a trade-off regarding which properties are better suited to adjust simple pairwise potentials. A more accurate description of

a transport property compromises the description of equilibrium properties. Furthermore, the melting pressure cannot be accurately modelled using the Mie potential if the fluid phase is to be accurately described. Presumably, the coarse-grained approximation overlooks structural effects that influence the conformation of the crystal phase, as well as the friction between molecules that affects the transport properties. These results indicate that the coarse-grained approximation is, in most cases, overfitted to a single property, and extrapolation to other properties must be approached with caution.

The results are presented for nitrogen in Fig. 3. Fig. 3(a)–(c) illustrate the VLE results using the fluid FE-ANN EoS. A good agreement can be observed for the VLE phase envelope, vaporisation pressure, and vaporisation enthalpy using OF_1 or OF_2. However, it is important to note that the fluid FE-ANN EoS is stretched to predict VLE at conditions below the triple temperature. In a molecular simulation context, the Mie particle exhibits coexistence between a fluid and a solid phase in those conditions. In Fig. 3(d)–(f), the entire phase diagram is modelled using the FE-ANN(s) EoS with different parameter sets. These figures clearly demonstrate that the triple point

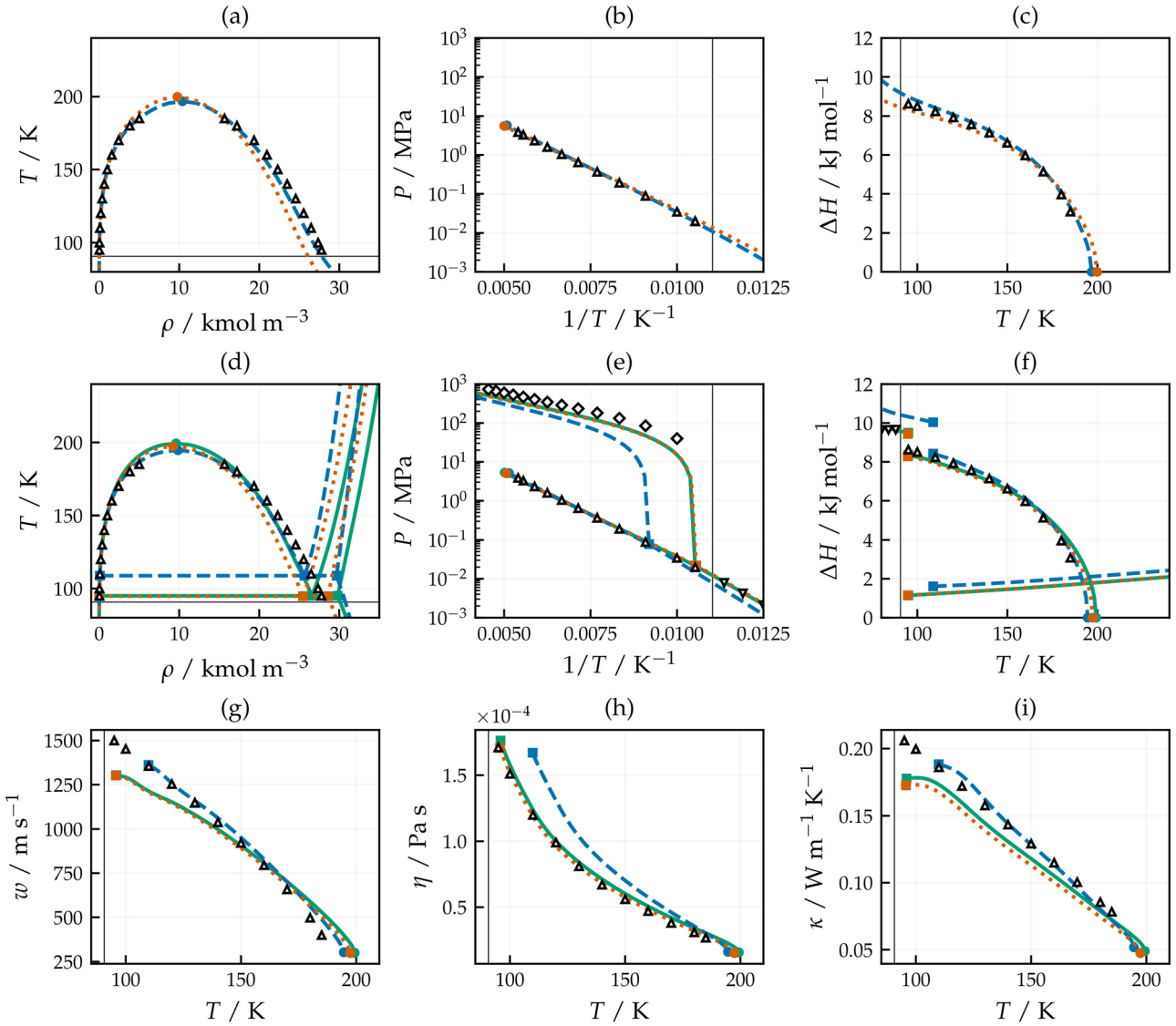


Fig. 4 Phase equilibria modelling of methane. (a) and (d) Phase envelope. (b) and (e) Clapeyron plot. (c) and (f) Enthalpy of phase change. (g) Speed of sound. (h) Shear viscosity. (i) Thermal conductivity. (a)–(c) Phase equilibria modelled using the FE-ANN EoS developed ref. 40. (d)–(g) Phase equilibria and speed of sound modelled using the FE-ANN(s) EoS developed in ref. 41. FE-ANN/FE-ANN(s) EoSs are based on eqn (3). (h) and (i) The shear viscosity and thermal conductivity are modelled using the ANN-based models developed ref. 42, based on eqn (6) and (12). Dashed blue lines and symbols refer to results using parameters optimised with OF_1 (VLE only). Dotted orange lines and symbols refer to results using parameters optimised with OF_2 (VLE + shear viscosity). Solid green lines and symbols refer to results using parameters optimised with OF_3 (VLE + SLE + SVE). Refer to eqn (2) and Table 1 for further details about the objective functions. Filled circle is the critical point. Filled square is the triple point. Pseudo-experimental data obtained from NIST TRC.[43] Upright triangles are VLE data, diamonds are SLE data, and downward triangles are SLE data. For reference, the solid black line refers to the triple temperature.

predicted by the parameters obtained from OF_1 is above some experimental VLE data points. Under those conditions, the VLE is unstable, and the Mie particle exhibits SVE. Incorporating the shear viscosity or solid phase data reduces the repulsive exponent and interaction energy of the Mie potential and, consequently, lowers the triple point. The parameters from OF_2 and OF_3 model the VLE similarly to those from OF_1, but they provide a better description of the triple point, SLE and SVE. It is noteworthy that the order of magnitude of the melting pressure ($\sim 10^3$ MPa) is considerably higher than that of the vaporisation ($\sim 10^{-1}$ MPa) and sublimation pressures ($\sim 10^{-3}$ MPa). Although the molecular parameters obtained from OF_3 predict the correct melting pressure trend, the reported relative errors are significant (~50%). Ultimately, the differences in the various molecular parameters result in noticeably distinct SLE densities. This information could be helpful in distinguishing optimal molecular parameters; however, this data is not available from the NIST TRC correlations.[43]

The liquid speed of sound results of Nitrogen are shown in Fig. 3(g). It is observed here that the three parameterisation approaches led to a similar description of this property. However, it should be noted that the described

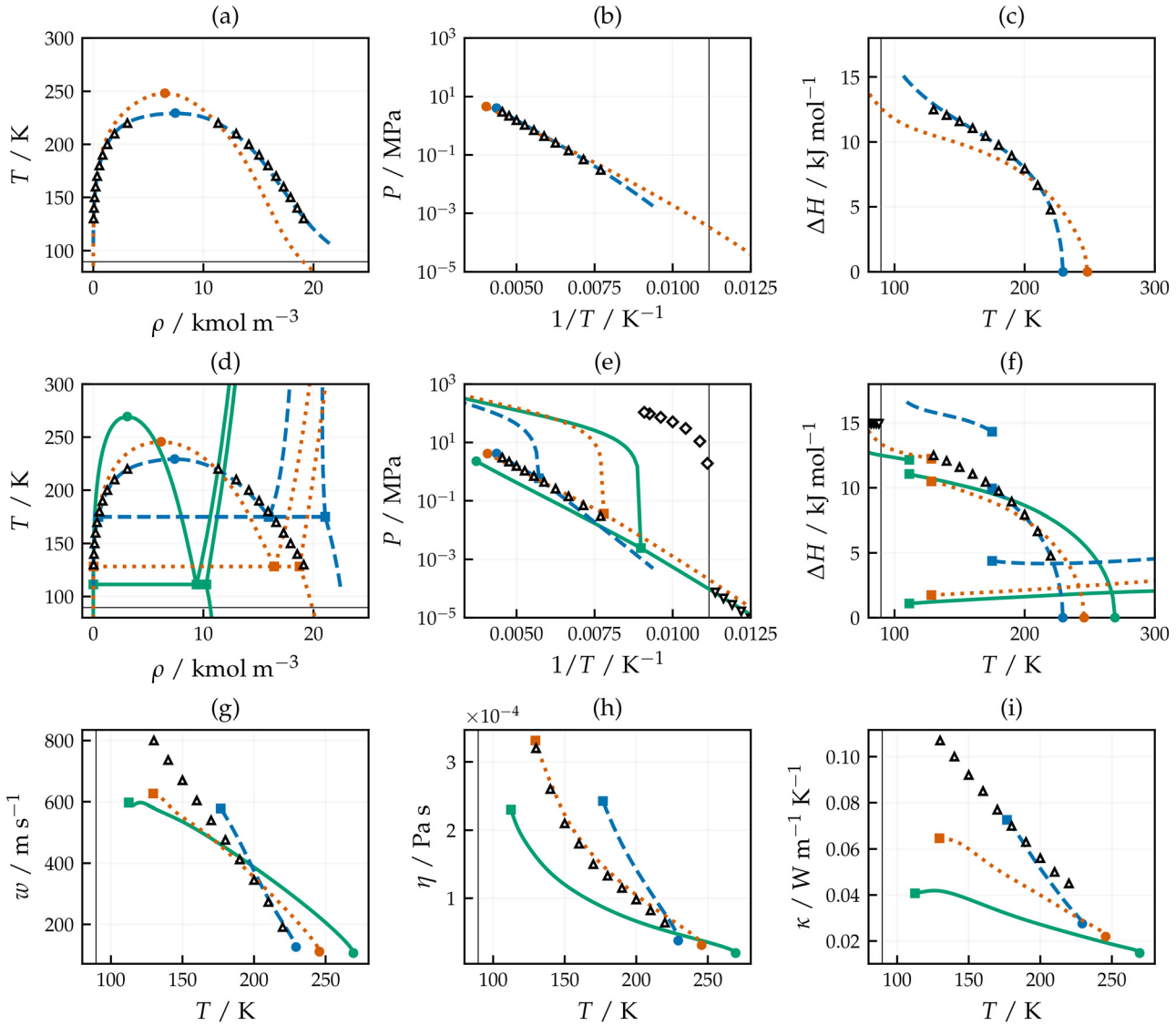


Fig. 5 Phase equilibria modelling of tetrafluoromethane. (a) and (d) Phase envelope. (b) and (e) Clapeyron plot. (c) and (f) Enthalpy of phase change. (g) Speed of sound. (h) Shear viscosity. (i) Thermal conductivity. (a)–(c) Phase equilibria modelled using the FE-ANN EoS developed ref. 40. (d)–(g) Phase equilibria and speed of sound modelled using the FE-ANN(s) EoS developed in ref. 41. FE-ANN/FE-ANN(s) EoSs are based on eqn (3). (h) and (i) The shear viscosity and thermal conductivity are modelled using the ANN-based models developed ref. 42, based on eqn (6) and (12). Dashed blue lines and symbols refer to results using parameters optimised with OF_1 (VLE only). Dotted orange lines and symbols refer to results using parameters optimised with OF_2 (VLE + shear viscosity). Solid green lines and symbols refer to results using parameters optimised with OF_3 (VLE + SLE + SVE). Refer to eqn (2) and Table 1 for further details about the objective functions. Filled circle is the critical point. Filled square is the triple point. Pseudo-experimental data obtained from NIST TRC.[43] Upright triangles are VLE data, diamonds are SLE data, and downward triangles are SLE data. For reference, the solid black line refers to the triple temperature.

ranges for these properties depend on the predicted critical and triple points. For the case of the shear viscosity, shown in Fig. 3(h), it can be observed that including properties related to the repulsive branch of the force field enhances the accurate description of this property. In this instance, the parameters from OF_2 and OF_3 exhibit a similar behaviour. This finding contrasts with the observations for the thermal conductivity, illustrated in Fig. 3(i). These results graphically illustrate the trade-offs inherent in the coarse-grained modelling approach, specifically regarding which properties to fit (in this case, shear viscosity or thermal conductivity). Similar conclusions can be drawn from the analysis of the data for carbon monoxide, which is included in section S.2 of the ESI.†

3.3 Tetrahedral molecules

This final section will analyse the limitations of the Mie potential in modelling tetrahedral molecules such as methane (CH_4) and tetrafluoromethane (CF_4). These molecules are often considered as isotropic spheres in UA or CG representations. The results for these two molecules are shown in Fig. 4 and 5, respectively. When only considering the VLE information using the fluid FE-ANN EoS, figures (a) to (c), there is a clear and excellent match to the phase envelope, saturation pressure, and vaporisation enthalpy. Results such as these have led to the claims that the Mie force field holds an advantage over the LJ model, in that the use of the additional parameter (λ_r) allows for a quantitative description of the fluid phase equilibria.[28] This is certainly the case, but the apparent success emerges from an overestimation of the repulsions and interaction energies of the force field. In the case of tetrafluoromethane, the optimisation programme reaches the upper limit of the FE-ANN EoS with $\lambda_r = 34$. The overfitting of the VLE properties is evident when analysing the high deviations of the shear viscosity and melting pressure in Table 3. As with the case of linear molecules, the use of overfitted parameters unphysically predicts VLE in regions below the triple point.

The effect of the various choices of parameters in the predicted phase diagram is shown in subfigures (d) to (i). For the case of Methane, Fig. 4, it observed that when including solid phase (OF_3) or shear viscosity (OF_2) data the correct behaviour for the VLE is retained with an improvement on the prediction of the triple point, SLE and SVE. Both OF_2 and OF_3 objective functions lead to lower repulsions and interaction energies of the system and similar molecular parameters. In this context, it is observed that incorporating a thermophysical property related to the repulsive interactions benefits the description of the phase diagram and transport properties. This result is also observed for linear molecules and aligns with the corresponding states methodology proposed by Hoang *et al.*[38]

For the case of tetrafluoromethane, Fig. 5, the limitations of the Mie potential are further unveiled. In Fig. 5(d)–(f), it is shown how the highly repulsive parameters (*i.e.*, $\lambda_r = 34$) obtained from OF_1 lead to a phase diagram where a considerable part of the experimental VLE data points are below the triple line. Below this temperature, the Mie particle will exhibit a solid phase. Moreover, when incorporating solid (OF_3) or shear viscosity (OF_2) data, there is no set of parameters that can adequately describe all properties (hence the title of this manuscript). It must be noted that all these parameter sets fail to describe the triple temperature of tetrafluoromethane. It is not a surprise that other thermophysical properties, such as the speed of sound, Fig. 5(g), and thermal conductivity, Fig. 5(i), are not described correctly either. Tetrafluoromethane (CF_4), similar to methane (CH_4), is a non-polar molecular. However, the fluorine atoms in CF_4 are highly electronegative and larger than the hydrogen atoms in CH_4. The fluoride atoms in CF_4 can lead to strong local polar moments, an overall octopolar moment and steric effects that a simple single-bead coarse-grained model seems incapable of taking into account. These results align with the findings of Bell[57] that suggested including these polar moments in an EoS to improve the description of derivative data of refrigerants.

4 Conclusions

This manuscript highlights the limitations of the Mie potential for modelling quasi-spherical molecules. The molecules studied include monoatomic (*e.g.*, argon, krypton, and xenon), linear (*e.g.*, nitrogen and carbon monoxide), and tetrahedral (*e.g.*, methane and tetrafluoromethane) molecules. While other pseudo-spherical molecules could have been included, the results obtained are to a large extent repetitive. These molecules were parameterised using three approaches. The first parameterisation approach relied solely on vapour–liquid equilibria data. The second parameterisation approach took into account the liquid shear viscosity and the vapour–liquid equilibria data. Finally, the third parameterisation approach included solid–fluid (solid–liquid and solid–vapour) equilibrium data in addition to the vapour–liquid equilibria data.

It has been found that a single set of parameters can accurately describe the entire range of properties for simple monoatomic molecules, such as argon, krypton, and xenon. An optimal set of parameters is encountered regardless of the thermophysical properties used to parameterise the Mie potential. The expected simple nature of noble gases, described as isotropic particles with a fixed repulsion and dispersion, most likely contributes to this result. However, the limitations of the Mie potential become evident beyond these simple cases. For linear and tetrahedral molecules, the numerical solution of the parameterisation overfits the vapour–liquid equilibrium when it is the sole target property. Notably, this approach is a customary parametrisation strategy in SAFT-like EoSs[25,39] and leads to a displacement of the triple point, resulting in a region of the VLE phase envelope that is unstable (frozen) as it lies below the triple temperature. This behaviour is ignored when using the EoS

in the context of fluid phases, but it will lead to incorrect phase behaviour when deployed in a molecular simulation. Furthermore, this suboptimal parameterisation leads to a poor description of other thermophysical properties, particularly transport properties. In contrast to noble gases, which interact solely through simple repulsion and van der Waals forces, polyatomic molecules can exhibit additional anisotropic inter- and intramolecular interactions. Examples include dipolar and higher polar moments, polarisation effects, and steric effects caused by molecular geometry. These interactions are not accounted for when modelling a polyatomic molecule using a coarse-grained approach with a single bead per molecule.

In contrast to the exclusive use of VLE data, the proposed parameterisation approaches suggest that incorporating properties such as shear viscosity and/or solid–fluid equilibria data results in a different set of molecular parameters. The inclusion of these properties in the parameterisation influences the interaction energy and repulsive exponent of the molecular model. This stems from the fact that both of these properties are strongly affected by short-range interactions. Shear viscosity relates to the friction between molecules, which is pertinent at short distances. Likewise, the solid phase and crystal structure directly relate to relatively short-distance steric packing. In contrast, vapour–liquid equilibria is a consequence of the interplay between repulsive and attractive interactions. This interplay can be characterised by multiple sets of roughly equivalent molecular parameters that correctly describe the VLE region.[19] This multiplicity (and ultimate mathematical global minimum in the parameter search) is not a guarantee that the parameters will extrapolate effectively to other thermophysical properties. For this reason, we advise to include properties related to the repulsions of the system when parameterising force fields. This observation aligns with the corresponding states methodology proposed by Hoang *et al.*,[38] which takes the shear viscosity into account in the force field parametrisation.

Incidentally, the use of viscosity or solid properties in the overall fit reduce the rather large repulsive exponent ($\lambda_r > 20$) obtained when "force-fitting" large or polar molecules to VLE. The apparent success of UA force fields based on the Mie potential in describing both fluid phase equilibria and transport properties,[58,59] despite being fitted only to VLE, seems to be aided by a fortunate choice of a conservative repulsive exponent within the range of 14–16. We conjecture that more robust and transferable parameter sets could be obtained by incorporating information such as viscosity, triple point data, or similar that directly pertains to the repulsive branch of the potential.

In summary, we have found that

- Relying exclusively on VLE as the target property in the objective function carries a risk of "overfitting" the Mie potential. This seemingly "optimal" parameterisation does not guarantee that other thermophysical properties can be accurately modelled or transferred. A clear sign of overfitting, as well as of an inadequate molecular model, is the shift of the triple point, which reduces the effective VLE region predicted by the model compared to experimental data.
- The parametrisation based exclusively on VLE data does not sufficiently quantify system repulsions. We propose a compromise parameterisation that we believe offers greater robustness than solely concentrating on VLE data. We recommend including a property that is significantly influenced by these short-range interactions, such as viscosity or solid-phase data, in the force field parameterisation. The limited effectiveness of VLE modelling in this context highlights the shortcomings of the molecular model or force field employed (which is the reasoning behind the manuscript's running title). Often, the Mie model's parameter set cannot simultaneously account for all properties.

As an end note, we highlight that the parameterisation approaches presented in this manuscript rely on physics-informed artificial neural networks to model thermophysical properties. These thermophysical property models are equations of state derived directly from discrete molecular simulation data and physics-inspired constraints employing machine learning techniques. They serve as reliable surrogates for molecular simulation results and represent promising paths for parameterising force fields using a top-down approach. It should be noted that the observations made in this manuscript are limited to simple spherical geometries and London-like dispersive interactions. Future work will focus on extending the framework to incorporate additional molecular geometries and interactions.

Data availability

Data and Python scripts to reproduce the results of this article are found in the following repository: **https://github.com/gustavochm/Square-Peg-Mie**.

Author contributions

Gustavo Chaparro: conceptualization (equal), data curation (lead), formal analysis (equal), investigation (lead), methodology (equal), software (lead), validation (equal), writing – original draft (lead) Erich A. Müller: conceptualization (lead), funding acquisition (lead), formal analysis (equal), investigation (equal), methodology (equal), supervision (lead), validation (equal), writing – review & editing (equal).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

G. C. acknowledges the support provided by the President's Ph.D. Scholarship (Imperial College London). Computations

were performed using the Imperial College Research Computing Service (DOI: **https://doi.org/10.14469/hpc/2232**).

Notes and references

1 G. C. Maitland, *Intermolecular forces: Their origin and determination*, Clarendon Pr, Oxford, 1981.
2 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.
3 M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.
4 A. Hemmen and J. Gross, *J. Phys. Chem. B*, 2015, **119**, 11695–11707.
5 M. J. Tillotson, N. I. Diamantonis, C. Buda, L. W. Bolton and E. A. Müller, *Phys. Chem. Chem. Phys.*, 2023, **25**, 12607–12628.
6 O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K.-R. Müller, *Chem. Rev.*, 2021, **121**, 10142–10186.
7 J. Behler and G. Csányi, *Eur. Phys. J. B*, 2021, **94**, 142.
8 J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.
9 C. Schran, F. L. Thiemann, P. Rowe, E. A. Müller, O. Marsalek and A. Michaelides, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2110077118.
10 V. L. Deringer, D. M. Proserpio, G. Csányi and C. J. Pickard, *Faraday Discuss.*, 2018, **211**, 45–59.
11 D. P. Kovács, C. V. D. Oord, J. Kucera, A. E. A. Allen, D. J. Cole, C. Ortner and G. Csányi, *J. Chem. Theory Comput.*, 2021, **17**, 7696–7711.
12 D. P. Kovács, I. Batatia, E. S. Arany and G. Csányi, *J. Chem. Phys.*, 2023, **159**, 044118.
13 E. A. Müller and G. Jackson, *Annu. Rev. Chem. Biomol. Eng.*, 2014, **5**, 405–427.
14 K. R. Hadley and C. McCabe, *J. Phys. Chem. B*, 2010, **114**, 4590–4599.
15 T. C. Moore, C. R. Iacovella and C. McCabe, *J. Chem. Phys.*, 2014, **140**, 224104.
16 T. G. A. Youngs, M. G. Del Pópolo and J. Kohanoff, *J. Phys. Chem. B*, 2006, **110**, 5697–5707.
17 J. Jin, A. J. Pak, A. E. P. Durumeric, T. D. Loose and G. A. Voth, *J. Chem. Theory Comput.*, 2022, **18**, 5759–5791.
18 F. L. Thiemann, P. Rowe, E. A. Müller and A. Michaelides, *J. Phys. Chem. C*, 2020, **124**, 22278–22290.
19 N. Ramrattan, C. Avendaño, E. Müller and A. Galindo, *Mol. Phys.*, 2015, **113**, 932–947.
20 F. London, *Trans. Faraday Soc.*, 1937, **33**, 8b.
21 J. Fischer and M. Wendland, *Fluid Phase Equilib.*, 2023, **573**, 113876.
22 J. E. Jones, *Proc. R. Soc. London, Ser. A*, 1924, **106**, 441–462.
23 J. E. Jones, *Proc. R. Soc. London, Ser. A*, 1924, **106**, 463–477.
24 T. M. Reed and K. E. Gubbins, *Applied statistical mechanics: thermodynamic and transport properties of fluids*, McGraw-Hill, 1973.
25 T. Lafitte, A. Apostolakou, C. Avendaño, A. Galindo, C. S. Adjiman, E. A. Müller and G. Jackson, *J. Chem. Phys.*, 2013, **139**, 154504.
26 S. Rahman, O. Lobanova, G. Jiménez-Serratos, C. Braga, V. Raptis, E. A. Müller, G. Jackson, C. Avendaño and A. Galindo, *J. Phys. Chem. B*, 2018, **122**, 9161–9177.
27 C. Avendaño, T. Lafitte, A. Galindo, C. S. Adjiman, G. Jackson and E. A. Müller, *J. Phys. Chem. B*, 2011, **115**, 11154–11169.
28 C. Avendaño, T. Lafitte, C. S. Adjiman, A. Galindo, E. A. Müller and G. Jackson, *J. Phys. Chem. B*, 2013, **117**, 2717–2733.
29 O. Lobanova, C. Avendaño, T. Lafitte, E. A. Müller and G. Jackson, *Mol. Phys.*, 2015, **113**, 1228–1249.
30 P. Morgado, O. Lobanova, E. A. Müller, G. Jackson, M. Almeida and E. J. M. Filipe, *Mol. Phys.*, 2016, **114**, 2597–2614.
31 C. Herdes, E. Forte, G. Jackson and E. A. Müller, *Adsorpt. Sci. Technol.*, 2016, **34**, 64–78.
32 G. Jiménez-Serratos, C. Herdes, A. J. Haslam, G. Jackson and E. A. Müller, *Macromolecules*, 2017, **50**, 4840–4853.
33 M. Fayaz-Torshizi and E. A. Müller, *Mol. Syst. Des. Eng.*, 2021, **6**, 594–608.
34 M. Fayaz-Torshizi and E. A. Müller, *Macromol. Theory Simul.*, 2022, **31**, 2100031.
35 A. Mejía, C. Herdes and E. A. Müller, *Ind. Eng. Chem. Res.*, 2014, **53**, 4131–4141.
36 C. Herdes, T. S. Totton and E. A. Müller, *Fluid Phase Equilib.*, 2015, **406**, 91–100.
37 Å. Ervik, A. Mejía and E. A. Müller, *J. Chem. Inf. Model.*, 2016, **56**, 1609–1614.
38 H. Hoang, S. Delage-Santacreu and G. Galliero, *Ind. Eng. Chem. Res.*, 2017, **56**, 9213–9226.
39 V. Papaioannou, T. Lafitte, C. Avendaño, C. S. Adjiman, G. Jackson, E. A. Müller and A. Galindo, *J. Chem. Phys.*, 2014, **140**, 054107.
40 G. Chaparro and E. A. Müller, *J. Chem. Phys.*, 2023, **158**, 184505.
41 G. Chaparro and E. A. Müller, *Commun. Phys.*, 2024, **7**, 406.
42 G. Chaparro and E. A. Müller, *J. Phys. Chem. B*, 2024, **128**, 551–566.
43 P. Linstrom, NIST Chemistry WebBook, NIST Standard Reference Database 69, 1997, http://webbook.nist.gov/chemistry/.
44 M. L. Michelsen and J. M. Mollerup, *Thermodynamic models: fundamentals & computational aspects*, Tie-Line Publications, Holte, 2nd edn, 2007.
45 D. B. Newell, F. Cabiati, J. Fischer, K. Fujii, S. G. Karshenboim, H. S. Margolis, E. De Mirandés, P. J. Mohr, F. Nez, K. Pachucki, T. J. Quinn, B. N. Taylor, M. Wang, B. M. Wood and Z. Zhang, *Metrologia*, 2018, **55**, L13–L16.
46 W. Wilding, R. L. Rowley and J. L. Oscarson, *Fluid Phase Equilib.*, 1998, **150–151**, 413–420.
47 G. Galliero and C. Boned, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2009, **80**, 061202.
48 B. E. Poling, J. M. Prausnitz and J. P. O'Connell, *Properties of Gases and Liquids*, McGraw-Hill Education, 5th edn, 2020.
49 Z. Liang and H.-L. Tsai, *Mol. Phys.*, 2010, **108**, 1707–1714.
50 F. Gao and L. Han, *Comput. Optim. Appl.*, 2012, **51**, 259–277.

51 P. Virtanen, *et al.*, *Nat. Methods*, 2020, **17**, 261–272.
52 I. Sobol', *USSR Comput. Math. Math. Phys.*, 1967, **7**, 86–112.
53 L. Zheng, J. M. Trusler, F. Bresme and E. A. Müller, *Fluid Phase Equilib.*, 2019, **496**, 1–6.
54 J. Šlepavičius, A. Patti and C. Avendaño, *Phys. Fluids*, 2025, **37**, 017129.
55 K. Meier and S. Kabelac, *Rev. Sci. Instrum.*, 2006, **77**, 123903.
56 C. Wedler and J. P. M. Trusler, *J. Chem. Eng. Data*, 2023, **68**, 1305–1312.
57 I. Bell, *Proceedings of the 20th International Refrigeration and Air Conditioning Conference*, West Lafayette, IN, United States, 2024.
58 R. A. Messerly, M. C. Anderson, S. M. Razavi and J. R. Elliott, *Fluid Phase Equilib.*, 2019, **483**, 101–115.
59 M. Fischer, G. Bauer and J. Gross, *Ind. Eng. Chem. Res.*, 2020, **59**, 8855–8869.