Results in Physics 52 (2023) 106782

Contents lists available at ScienceDirect

Results in Physics

journal homepage: www.elsevier.com/locate/rinp

Determination of pairwise interactions via the radial distribution function in equilibrium systems interacting with the Mie potential

Jianxiang Tian [a,b,c,*], Ludovic Berthier [d,e]

[a] Department of Physics, Qufu Normal University, Qufu 273165, PR China
[b] Department of Physics, Dalian University of Technology, Dalian 116024, PR China
[c] XihouYougu Institue for Advanced Study, Sishui 273200, PR China
[d] Laboratoire Charles Coulomb (L2C), Universite de Montpellier, CNRS, 34095 Montpellier, France
[e] Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

ARTICLE INFO

Keywords:
Interaction potential
Radial distribution function
Molecular dynamical simulations

ABSTRACT

Particle interactions play a fundamental role in condensed matter physics because they determine both the dynamical behavior and the equilibrium physical properties of a given system at temperature T and density ρ. However, these interactions are not always precisely known in experiments, or in simulations of coarse-grained systems. A direct determination of the pair interaction potential in a given system could help understand observed behaviors and make further predictions. Given a number of equilibrium configurations of a system, it would be desirable to find a method to directly determine the pair potential by only using these snapshots. We propose two simple methods towards this goal for the specific case of the systems in 3 dimensional space with the Mie potential, which includes two exponents as m and s. Well-equilibrated system configurations are produced by molecular dynamical simulations using the Mie potential with different exponent combinations (m, s). In the first method, we construct a correspondence between the value and location of the first peak of the radial distribution function and the couple (m, s), which allows us to determine the potential with an accuracy of 100% when given a set of equilibrium configurations for an unknown potential. In the second method, we train an artificial neural network to learn this correspondence. We find that all (m, s) combinations are correctly predicted. Both methods support the idea that the pairwise interaction can often easily be inferred by using equilibrium snapshots.

Introduction

Our world is colorful due to the existence of interactions among particles. Particle–particle interactions are of paramount importance in many-body systems as they determine the collective behavior and coupling strength [1]. For instance, in the simple case of the hard sphere interaction model, systems made of hard spheres display gas [2–4], fluid [3–8], crystals [9–14], disordered glasses [12,15] as well as hyperuniform structures [16]. For any model system, one can obtain all of its physical properties, including its phase diagram and microstructure, when the interactions among particles are fixed. In other words, interactions play a fundamental role in physics.

Often, particle–particle interactions are too complex to be obtained accurately in real physical systems. Very complicated expressions describing such interactions can be found in material science and chemical physics, obtained by different types of experimental and numerical methods [17–20]. In experiments, the particle–particle interactions can be completely unknown, while a determination of effective pairwise interactions can be useful to describe complex systems at a more coarse-grained level in computer simulations [21], and can also be useful for determining experimental parameters such as particle charge or screening length in colloidal systems, assuming the interaction is the Yukawa potential. Another potential application is the determination of the optimal pair potential to be used in order to produce a desired targeted structure, as in self-assembly [22].

In practice, one has instead access to some dynamical behavior or to some physical structural properties at equilibrium at the corresponding experimental conditions. Only when the particle–particle interactions in a system are well-characterized can one understand better its observed behaviors and even make predictions about unknown behaviors in different conditions. Thus, it is of great importance to find ways to determine particle–particle interactions directly from observations [23,24].

Conceptually, the pairwise interaction for a specific physical system can be determined by analyzing the statistics of a system comprised of

* Corresponding author at: Department of Physics, Qufu Normal University, Qufu 273165, PR China.
E-mail addresses: jxtian@qfnu.edu.cn, jxtian@dlut.edu.cn (J. Tian).

https://doi.org/10.1016/j.rinp.2023.106782
Received 5 April 2023; Received in revised form 13 July 2023; Accepted 20 July 2023
Available online 23 July 2023
2211-3797/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

only two particles [25]. Although direct and simple, this method is not always accessible (for instance in experimental work or when a coarse-grained potential is sought) and it is desirable to devise methods that extract the pair interaction directly from the many-body interacting system at the desired thermodynamic conditions. In this context, this amounts to solving the inverse problem of extracting the pair potential from a given set of produced data. It is well known that, given the pair correlation function $g(r)$ there exists a one-to-one correspondence to the pair interaction potential $u(r)$ in virtue of the theorem derived by Henserson [26]. In practice however, there exists no analytic solution to this problem which must be tackled numerically. The most common method is perhaps the iterative Boltzmann inversion [27] where a candidate pair potential is proposed, and a computer simulation is used to measure its corresponding $g(r)$. The distance to the target pair correlation is used to improve the candidate potential in an iterative manner, in a way which can be refined [28]. Because an entire simulation is needed at each step the method is rather costly, and simpler methods would be welcome. Recently, an iterative method which does not require multiple simulations has been devised based in the particle insertion method [29]. This is an interesting step, but it only works for system that are dilute enough that particles can be inserted, which limits its applicability to dense liquids or solid phases.

In this paper, we focus on the systems interacting with the Mie potential in 3 dimensional space and devise a very simple idea to determine the two exponents of the Mie potential by only using an ensemble of equilibrium configurations obtained at thermal equilibrium. Although this problem is simpler than a full inversion problem to determine a completely unknown potential, it remains non-trivial and impossible to tackle analytically. Proposing a simple solution to the inverse problem in that narrower case is therefore a valuable contribution.

The manuscript is organized as follows. In Section "Theoretical basis", we explain the theoretical basis for the idea. In Section "Molecular dynamics simulations and numerical computations by Method 1", we provide details of our molecular dynamical simulations and the method of fitting correlations to check whether the idea well stands or not. In Section "Numerical computations by Method 2: artificial neural network (ANN)", we provide the second method of the artificial neural network model underneath the Machine Learning to check whether the idea well stands or not. In Section "Results and discussion", we demonstrate the validity of the predicted potentials under these two methods. In Section "Conclusion", we make a short conclusion.

Theoretical basis

As is well-known, a pairwise additive interaction potential is defined as a mathematical function of the distance between the two particles considered. Many different kinds of pairwise potential functions are being used in the literature, such as the hard sphere potential [23], the Lennard-Jones potential [30] which is a special case of the more general Mie potential [31], or the Yukawa potential with or without hard core [29]. The Mie potential reads [31]

$$u\left(r_{ij}\right)=\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{m}-\left(\frac{\sigma}{r_{ij}}\right)^{s}\right]. \tag{1}$$

When $m = 12$ and $s = 6$, Eq. (1) reduces to the Lennard-Jones potential [31]. In this paper, we consider equilibrated systems characterized by the Mie potential with 51 different (m, s) combinations in 3 dimensional space.

The radial distribution function (RDF) (also called pair correlation function) reads [32]

$$g_2\left(\boldsymbol{r}_1,\boldsymbol{r}_2\right)=\frac{V^2N!}{N^2\left(N-2\right)!}\frac{1}{Z_N}\int\cdots\int e^{-\beta U_N}d\boldsymbol{r}_3\cdots d\boldsymbol{r}_N. \tag{2}$$

Here $\{\boldsymbol{r}_i\}$ are the particle coordinates, $U_N\left(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N\right)$ is the potential energy of the system due to the interactions between particles, $Z_N = \int\cdots\int e^{-\beta U_N}d\boldsymbol{r}_1\cdots d\boldsymbol{r}_N$ is the configurational integral taken over all possible combinations of particle positions in the canonical ensemble. It is related to the canonical partition function Z as

$$Z_N = Z\cdot N!\lambda^{3N}, \tag{3}$$

where $\lambda=h\left(\frac{\beta}{2\pi m}\right)^{1/2}$ is the de Broglie thermal wavelength.

If the fluid system consists of spherically symmetric particles and the potential is fixed, $g_2\left(\boldsymbol{r}_1,\boldsymbol{r}_2\right)$ depends only on the relative distance between particles 1 and 2, $r_{12}=|\boldsymbol{r}_2-\boldsymbol{r}_1|$, so that

$$g_2\left(\boldsymbol{r}_1,\boldsymbol{r}_2\right)=g_2\left(r_{12}\right). \tag{4}$$

The canonical partition function for the fluid system composed of spherically symmetric particles with the Mie potential can be obtained using mean-field methods [33]

$$Z=\frac{V^N}{N!\lambda^{3N}}\exp\left(-A\rho^{\frac{m}{3}-1}+W\rho^{\frac{s}{3}-1}\right), \tag{5}$$

with $A=\frac{2\pi\epsilon\sigma^m}{m-3}\left(\frac{3}{4\pi}\right)^{1-m/3}$, $W=\frac{2\pi\epsilon\sigma^s}{s-3}\left(\frac{3}{4\pi}\right)^{1-s/3}$ and $\rho = N/V$. As a result, we have that,

$$Z=Z\left(\rho,\beta,m,s\right). \tag{6}$$

Combining Eqs. (2), (4), (6), the radial distribution function can be written as

$$g_2=g_2\left(r_{12},\rho,\beta,m,s\right) \tag{7}$$

Eq. (7) is reasonable as when the Mie potential is fixed (and so (m, s) are fixed) and the system is equilibrated at constant density ρ and temperature $T = 1/\beta$, we get that g_2 is a unique function of r_{12}. For a system with fixed density and temperature,

$$g_2=g_2\left(r_{12},m,s\right)=g_2\left(r,m,s\right). \tag{8}$$

As is also well-known, for a fluid system at equilibrium, the radial distribution function exhibits several peaks. We characterize the first peak by its location and amplitude $\left(r_{1st},g_{1st}\right)$. By solving

$$\begin{cases} g_{1st}=g_2\left(r_{1st},m,s\right) \\ \left.\dfrac{\partial g_2}{\partial r}\right|_{r=r_{1st}}=0 \end{cases} \tag{9}$$

one can then obtain m and s because r_{1st} and g_{1st} are unique functions of (m, s).

The above reasoning proposes an explicit demonstration, within a mean-field approach, that the exponents m and s can be determined directly from the pair correlation function. Given its approximate nature, we cannot use these expressions to perform the inversion in Eq. (9) analytically and we must thus resort to computer simulations. If the values of r_{1st} and g_{1st} are measured accurately, one can obtain the values of m and s by inverting the expressions for r_{1st} and g_{1st}. It is interesting to discuss the physical meaning of some special points of the RDF, for instance, the peaks. Peaks in the RDF reflect some important physical microscopic structures. As qualitatively deduced by Liu et al. [34], a structure with higher symmetry will produce a lower density of RDF peaks. One can capture at least three independent pieces of information in a peak of the RDF peak [35]: its position gives the average separation of the pair of atoms in question; its integrated intensity yields the coordination number of that pair of atoms; and the width and shape of the peak gives the underlying atomic probability distribution.

In the following, we use the distance histogram (DH) method to measure the values of r_{1st} and g_{1st} from a given set of data. The DH method is the most widely used method to measure the radial distribution function as a function of the distance between the two particles [36,37]. This is the basis for the numerical analysis proposed in the following sections.

Molecular dynamics simulations and numerical computations by Method 1

We define the reduced distances $r^* = r/\sigma$, density $\rho^* = \rho\sigma^3$, and temperature $T^* = k_B T/\epsilon$. In Section "Theoretical basis", we described the theoretical basis to obtain the two unknown exponents (m, s) from the value and location of the first peak of the RDF of equilibrated system configurations obtained with the Mie potential.

To test this idea, we perform computer simulations. We simulate Mie potential systems composed of $N = 2048$ particles at six different reduced densities of $\rho^* = 0.8$, 0.9, 1.0, 1.1, 1.2, and 1.3 and three reduced temperature of $T^* = 1/1.30, 1/1.35, 1/1.40$ using conventional molecular dynamics simulations in 3 dimensional space. Temperature is imposed using a Nose–Hoover thermostat. Periodic boundary conditions are used. Specially, the reduced time step reads $\Delta t^* = 0.005(t^* = t\sqrt{\epsilon/m}/\sigma)$, the reduced maximal time reads $t^*_{max} = 500$, and the inertia parameter reads 10. The analysis and discussion in the present work are based on the converged results after 10^5 time steps using a cutoff radius of $r_{cut} = L/2$ with L being the linear size of the simulated box to guarantee that all systems are equilibrated. The core Fortran codes are from Frenkel and Smith [24] and available on request. The exponents of the Mie potential chosen in the interval $m \in [8, 13]$ and $s \in [2, m-1]$. All systems were found to be thermodynamically stable. Note that the global stability of the considered cases of $s = 2$ and $s = 3$ confirmed numerically is presumably an artifact resulting from the chosen r_{cut} value, and that these systems would be thermodynamically unstable when considering infinite-range interactions. In total, we have studied 51 (m, s) combinations. For each equilibrated Mie system with fixed (m, s), 10^4 statistically independent configurations are stored to compute the radial distribution function using the DH method using a regular grid with discretization $\Delta r = 0.01\sigma$. From these numerical results, we obtain $g_2(r)$, and we can extract the value and location of its first peak (r_{1st}, g_{1st}) for each (m, s) combination. A detailed description of the DH method can be found elsewhere [33]. Numerically, we fit a short interval around the first peak using a third-order polynomial function to get a smooth curve and extract the value and location of its first peak from the corresponding polynomial function.

The measured values for the combinations $(m = 8, s = 2, 3, 4, 5, 6, 7)$, $(m = 10, s = 2, 3, 4, 5, 6, 7, 8, 9)$ $(m = 12, s = 2, 4, 6, 7, 8, 10)$ and $(m = 13, s = 2, 4, 6, 8, 10, 12)$ provide a total of 26 data points which are used to obtain empirical expressions for (r_{1st}, g_{1st}). For instance, in the case of $\rho^* = 0.8$ and $T^* = 1/1.35$ the fitting forms read:

$$\begin{aligned} r_{\rm fit}(m,s) =& 0.7044 + 0.06615m + 0.1906s - 0.003014m^2 - 0.03709ms \\ & - 0.03859s^2 + 0.001775m^2 s + 0.007326ms^2 + 0.001251s^3 \\ & - 0.0003539m^2 s^2 - 0.0002707ms^3 - 1.944\times 10^{-6} s^4 \\ & + 1.45\times 10^{-5} m^2 s^3 + 9.892\times 10^{-7} ms^4 - 1.103\times 10^{-6} s^5, \end{aligned} \tag{10}$$

$$\begin{aligned} g_{\rm fit}(m,s) =& 0.6661 + 0.2141m + 0.08667s - 0.00694m^2 - 0.04018ms \\ & - 0.01882s^2 + 0.002458m^2 s + 0.008879ms^2 - 0.003937s^3 \\ & - 0.0005022m^2 s^2 - 4.325\times 10^{-5} ms^3 + 0.0001732s^4 \\ & + 7.691\times 10^{-6} m^2 s^3 + 1.102\times 10^{-5} ms^4 - 1.2\times 10^{-5} s^5. \end{aligned} \tag{11}$$

Equivalent expressions for the other five densities are enclosed for the case of $T^* = 1/1.35$ in the Supporting Material (SM). Other expressions for other two cases of $T^* = 1/1.30, 1/1.40$ have also been obtained, see SM. For each density, we simulate a total of 51 combinations of exponents. We checked that the quality of the predictions became less if we only use lower-order polynomial in Eqs. (10), (11).

To check the accuracy of a given prediction for $(r_{\rm fit}, g_{\rm fit})$, we introduce an error function between the prediction and the measured values defined as

$$\chi^2 = \left(\frac{g_{\rm fit} - g_{1st}}{g_{1st}}\right)^2 + \left(\frac{r_{\rm fit} - r_{1st}}{r_{1st}}\right)^2. \tag{12}$$

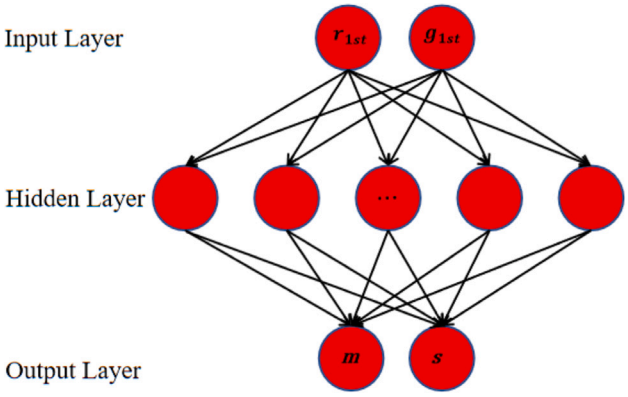

Fig. 1. Schematic structure of the artificial neural network. The inputs are (r_{1st}, g_{1st}) and the outputs are (m, s).

With these definitions, our procedure to determine the unknown exponents (m, s) from a given set of equilibrium configurations is as follows:

(*a*) For a set of 10^4 configurations for which we wish to infer the exponents (m, s) we measure numerically $g_2(r)$ by the DH method;

(*b*) We fit the region of the first peak using a polynomial function and extract the value and the location of the first peak (r_{1st}, g_{1st});

(*c*) For all possible pair (m, s) with integer values we use Eqs.(10), (11) to calculate $(r_{\rm fit}, g_{\rm fit})$;

(*d*) For each pair (m, s) we estimate the error χ^2 in Eq. (12) between the fitted and measured values;

(*e*) The (m, s) combination with the lowest χ^2 is selected as the best candidate for the two exponents of the Mie potential corresponding to the numerical data.

Numerical computations by Method 2: artificial neural network (ANN)

The spirit of machine learning (ML) is to automatically learn a set of rules from a set of data using an appropriate algorithm, to be able to make predictions for a new set of data. Artificial neural network (ANN) models act as a nonlinear data modeling device in ML [38–41]. We propose to use a very simple ANN model to determine the two exponents (m, s) of the Mie potential by using the value and location of the first peak of the RDF as inputs.

The structure of our ANN model is displayed in Fig. 1. The Matlab software is used to build the model. The ANN model is composed of the input layer, the hidden layer, and the output layer while the neurons exist in the layers. In the input layer, the input information is received, and each neuron represents an input variable and directly transmits the information to the neurons in the hidden layer without processing. In the hidden and output layers, each output receives a weighted sum of the inputs plus a bias, and then transmits the sum through an activation function to generate an output. We use a feed-forward neural network. The input variables are (r_{1st}, g_{1st}) and the output variables are the two exponents (m, s). Because the two exponents (m, s) to be determined are integers, we need to transform the real numbers obtained in output back into integers. We consider the case of $T^* = 1/1.35$ as an example to illustrate the details. For each density, the 51 (m, s) samples are divided into two disjoint sets of data. The numerical values of $(m = 8, s = 2, 3, 4, 5, 6, 7)$, $(m = 10, s = 2, 3, 4, 5, 6, 7, 8, 9)$, $(m = 12, s = 2, 4, 6, 7, 8, 10)$, $(m = 13, s = 2, 4, 6, 8, 10, 12)$, provide a total of 26 samples which form the part I (**PI**). The other 25 samples form the part II (**PII**). In **PI**, 80% of the 26 samples (20 samples) are used for training, 10% (3 samples) are used for validation, and 10% (3 samples) are used to perform the testing. These three databases are randomly assembled. Both 'tansig'

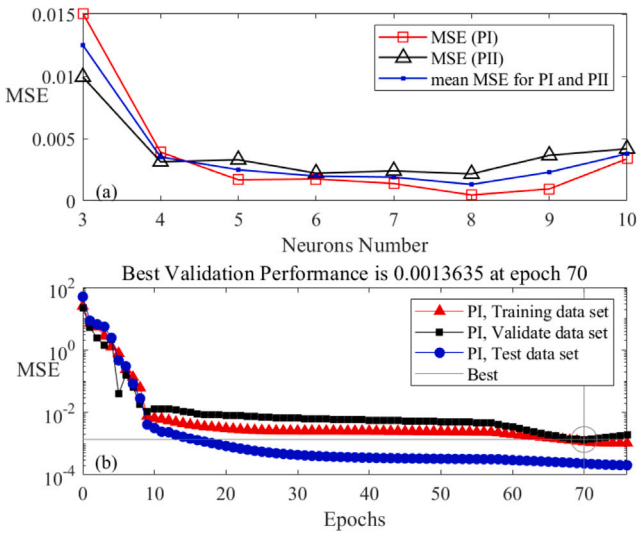

Fig. 2. (a) Mean-squared error of the artificial neural network for different numbers of neurons in the hidden layer. (b) Evolution of the mean-squared error with the number of epochs for $T^* = 1/1.35$ and $\rho^* = 0.8$ during the training of the artificial neural network.

and 'purelin' functions in Matlab software are selected as the activation functions in the hidden layer and output layer, respectively. The training accuracy of the ANN model is measured by the mean-squared error (MSE) defined as

$$\mathrm{MSE} = \frac{\sum(m_p - m_e)^2 + \sum(s_p - s_e)^2}{2N_s} \tag{13}$$

where (m_p, s_p) and (m_e, s_e) are respectively the predicted and exact values, and N_s is the number of samples. Using trial and error, we determined the optimal number of the neurons of the hidden layer with the mean square error (MSE), as is illustrated in Fig. 2(a). It is easy to see that the MSE is equal to 0.0004695 for **PI**, 0.002162 for **PII**, and 0.001316 for the mean MSE for 8 neurons.

The trained ANN model will be used for the prediction of the other 25 samples in part II (**PII**) which were not used in the model training stage. The other two temperature cases of $T^* = 1/1.30$ and $T^* = 1/1.40$ follow the same scheme but require different (m, s) combinations for **PI** and are available on request.

Results and discussion

Stability of the first peak of the RDF from the DH method

For a homogeneous system, the radial distribution function can be defined as [29] $g(\boldsymbol{r}) = \rho^*(\boldsymbol{r})/\rho$. It is given by the ratio of the local number density measured from a reference particle $\rho^*(\boldsymbol{r})$ over the bulk number density ρ. According to this definition, the RDF is typically measured using the distance histogram (DH) method [42].

We have followed the behavior of the measured value and location of the first peak of the RDF of the Lennard-Jones fluids at $\rho^* = 0.8$ and $T^* = 1/1.35$. For a small number of configurations (a few hundreds), the measured values appear to fluctuate due to statistical uncertainties. Only when the number of configurations is large enough the measured values become stable. We find that r_{1st} and g_{1st} become statistically reliable when the number of configurations becomes larger than $\approx 10^3$. We conclude that the measured g_{1st} will not take a constant value unless the number of configurations is quite large. In this work, the number of configurations (NC) is taken to be 10^4 for each (m, s) combination, which provides a compromise between numerical precision and computational load. According to the above discussion, NC = 10^4 is large enough to obtain a reasonable stability of both the value and the location of the first peak of the RDF. It is observed that both the obtained r_{1st} and g_{1st} by the DH method are completely stable with NC = 10^4 for all the (m, s) combinations we considered as follows. In the following, we show and analyze the results for the case of $T^* = 1/1.35$. Similar results hold for the other two temperature cases, as shown in Supporting Materials.

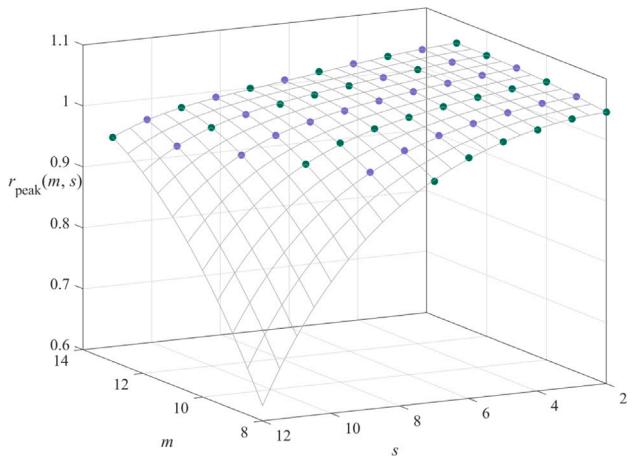

Fig. 3. Evolution of r_{1st} with m and s for $T^* = 1/1.35$ and $\rho^* = 0.8$. The surface represents Eq. (10). The dotted green symbols are the values used to fit Eq. (10). The dotted purple symbols are used for prediction. Similar curves for the other two temperature cases of $T^* = 1/1.30$ and $T^* = 1/1.40$ are shown in the Supporting Materials. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

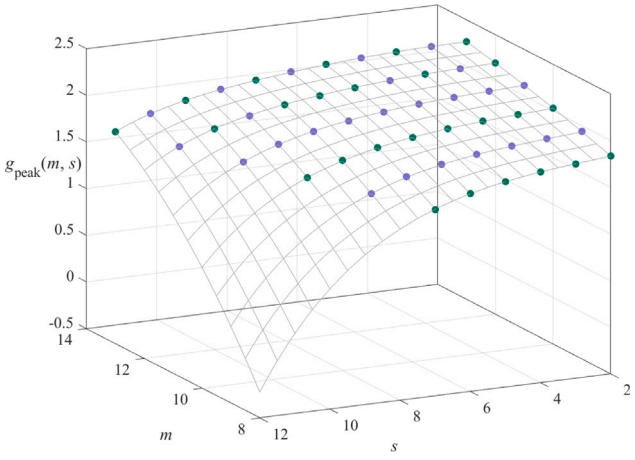

Fig. 4. Evolution of g_{1st} with m and s for $T^* = 1/1.35$ and $\rho^* = 0.8$. The surface represents Eq. (11). The dotted green symbols are used to fit Eq. (11). The dotted purple symbols are used for prediction. Similar curves for the other two temperature cases of $T^* = 1/1.30$ and $T^* = 1/1.40$ are shown in the Supporting Materials. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Numerical results for Method 1: Empirical fits

Equilibrated Mie systems with 51 different (m, s) combinations at $\rho^* = 0.8$, 0.9, 1.0, 1.1, 1.2, 1.3 and $T^* = 1/1.35$ are produced by MD simulations. The numerical results for the evolution of r_{1st} and g_{1st} with (m, s) are displayed in Figs. 3, 4. The curved surface in Fig. 3 represents Eq. (10), and the curved surface in Fig. 4 represents Eq. (11) while the symbols correspond to the measurements.

From Fig. 3, we see that the location of the first peak of the RDF decreases with s when m is fixed. Recall that m is the exponent of the repulsive part of the Mie potential and s is the exponent of the

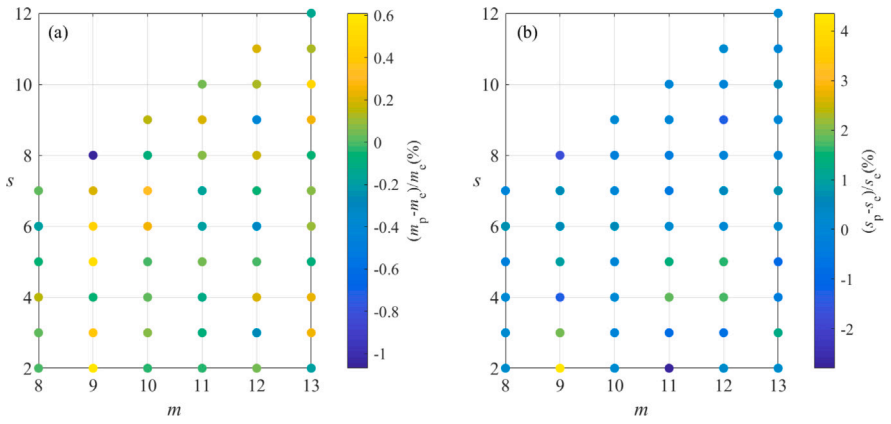

Fig. 5. Percentage deviation of the predictions made by the artificial neural network for each sample when $\rho^* = 0.8$ and $T^* = 1/1.35$ for decimal output. Panels show results for (a) m and (b) s predictions, respectively. When transformed into integer values, 100% of the predicted (m, s) values are correct.

Table 1
Number of (m, s) combinations in different R ranges. Here R is defined as $R = \chi_2^2/\chi_1^2$ with χ_1 the error for the 1st best candidate, χ_2 for the 2nd best candidate, and the error χ is defined in Eq. (12). Results for the other two temperature cases of $T^* = 1/1.30$ and $T^* = 1/1.40$ are shown in the Supporting Materials.

ρ^*	$R > 2$	$R > 4$	All
0.8	51	51	51
0.9	51	51	51
1.0	51	51	51
1.1	51	50	51
1.2	51	48	51
1.3	50	47	51

attractive part. Physically, increasing s at fixed m implies the increasing of attraction in large r and the decreasing of repulsion in small r, which enhances the first peak of the RDF and shifts its position to smaller r values.

We have a total of 51 different (m, s) combinations at each thermodynamic condition (ρ^*, T^*). We use 26 of them to obtain Eqs. (10), (11) and the other 25 are used to check the accuracy of the method proposed in Section "Molecular dynamics simulations and numerical computations by Method 1".

We find that all of the 25 (m, s) combinations that are not used to construct the fits at each thermodynamic condition (ρ^*, T^*) can be correctly predicted with no exception. We check that the 26 (m, s) combinations used for fitting can also be correctly predicted in the same manner.

In order to clearly show the difference between the 1st best candidate and the 2nd best candidate in each case, we have presented the results in Table 1, in which the number of combinations in different $R = \chi_2^2/\chi_1^2$ ranges are established, where χ_1 is the error obtained for the 1st best candidate and χ_2 the one for the 2nd best candidate, with the definition of χ in Eq. (12). For each density, $R > 2$ holds for more than 50 (m, s) combinations, and $R > 4$ for more than 47 combinations. Thus, the 1st candidate can be easily distinguished from the 2nd candidate.

In short, the correct (m, s) combination can be easily obtained by simply fitting two main characteristic values of the RDF with an accuracy of 100%.

Numerical results for Method 2: Artificial neural network

We choose the case $\rho^* = 0.8$ and $T^* = 1/1.35$ to present numerical results from the ANN model. The 51 (m, s) samples are divided into the **PI** and **PII** sets described in "Numerical computations by Method 2: artificial neural network (ANN)". It is worth noting that the MSE in Eq. (13) decreases and reaches a nearly constant value during the training. If the above condition is met, the ANN model can be expected to make good predictions.

Fig. 2(b) shows the trend of MSE with the number of epochs. The best validation performance is 0.0013635 at epoch 70 where the training performance is 0.0011548 and the test performance is 0.0002326. The three curves show a trend towards a nearly steady-state value.

Compared with the MSE, the percentage deviation can better reflect the prediction ability of ANN for individual combinations. Fig. 5 shows the percentage deviation (PD) achieved by the ANN model for all combinations of exponents when we let the outputs take decimal values. For m predictions shown in the panel (a), the absolute percentage deviation of all data points is less than $\approx$ 1%. The purple dot represents the maximal absolute percentage deviation value 1.07% at $(m = 9, s = 8)$. For the s predictions shown in the panel (b), the absolute percentage deviation of all data points is less than 5%. The yellow dot represents the maximal absolute percentage deviation value 4.35% at $(m = 9, s = 2)$. *When we transform the outputs into integer numbers, 100% of the predictions are correct.*

In other words, all 51 combinations in the case $\rho^* = 0.8, T^* = 1/1.35$ can be obtained by the ANN model. We add that the predictions are 100% correct without exceptions for the other five densities when we transform the outputs into integer numbers. The results for other densities are enclosed in the Supporting Material.

In summary, by using half of the data to train the ANN model, the two exponents of the Mie potential for all newly produced data can be correctly determined by simply analyzing the first peak of the radial distribution function.

Conclusion

In this paper, we established a simple connection between the characteristics of the first peak of the radial distribution function and the exponents of the Mie potential interaction for well-equilibrated configurations. Given a number of snapshots taken from an equilibrated ensemble created via the Mie potential with unknown exponents, we can directly determine the values of the exponents (m, s) with excellent accuracy. This work offers two simple ways to determine the pairwise interaction by only using configuration snapshots, which are easily obtained in numerical simulations or experiments. Therefore, our strategy considerably simplifies the general inverse problem where a determination of the pair potential from simulation data is needed, although it of course does not offer a complete general solution.

In conclusion, we have validated a simple idea to measure pairwise potentials. This idea can be implemented with high accuracy by either using both the simple method based on empirical fitting or using the artificial neural network method. In Method 1, we construct a correspondence between the value and location of the first peak of the radial distribution function with the couple (m, s), which allows us to then determine the potential with an accuracy of 100% when given

a set of equilibrium configurations for a new system with unknown potential. In Method 2, an artificial neural network is trained at each density by using half of the samples, and the remaining samples are correctly predicted with an accuracy of 100% by the model, when the predicted (m, s) are transformed into integer values.

Both methods support the idea that easy and directly determinations of the pairwise interaction can be obtained by only using equilibrium snapshots. This idea can be readily extended to multicomponent systems, and will be valuable in both fundamental and applied research. This idea can also be applied to fundamental investigations of many-body interactions, such as coarse-grained pair potentials [27,43,44] in multiscale simulations of complex fluids and biological systems and polymer systems. The extension to any possible pairwise potential interaction such as the Yukawa potential are in progress.

CRediT authorship contribution statement

Jianxiang Tian: MD simulations, Data curation, Writing – original draft, Writing – review & editing. **Ludovic Berthier:** MD simulations, Data curation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request

Acknowledgments

The National Natural Science Foundation of China under Grant No. 11274200 (J. T.), the Natural Science Foundation of Shandong Province, China under Grant No. ZR2022MA055 (J. T.), the foundations with Grant No. 2022NS189 from XYIAS (J. T.), and the Simons Foundation, United States (#454935 L. B.) have supported this work. Tian sincerely thanks Dr. Q. Chen and Y. Liu for their help. The authors would like to show their special thanks to the anonymous referee for his/her helpful comments and suggestions.

Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.rinp.2023.106782.

References

[1] Misiunas K, Pagliara S, Lauga E, Lister JR, Keyser UF. Nondecaying hydrodynamic interactions along narrow channels. Phys Rev Lett 2015;115:038301.
[2] Gust ED, Reichl LE. Molecular dynamics simulation of collision operator eigenvalues. Phys Rev E 2009;79:031202.
[3] Mulero A. Lecture notes in physics, Vol. 753. Berlin, Heidelberg: Springer; 2008.
[4] Santos A. Lecture notes in physics, Vol.923. Berlin, Heidelberg: Springer; 2016.
[5] Kolafa J, Labik S, Malijevsky A. Accurate equation of state of the hard sphere fluid in stable and metastable regions. Phys Chem Chem Phys 2004;6:2335.
[6] Tian JX, Jiang H, Mulero A. Equations of the state of hard sphere fluids based on recent accurate virial coefficients B5–B12. Phys Chem Chem Phys 2019;21:13070.
[7] Santos A, Yuste SB, Lòpez de Haro M. Structural and thermodynamic properties of hard-sphere fluids. J Chem Phys 2020;153:120901.
[8] Hu JW, Yu YX. High-order virial coefficients and equation of state for hard sphere and hard disk systems. Phys Chem Chem Phys 2009;11:9382.
[9] Pronk S, Frenkel D. Large difference in the elastic properties of fcc and hcp hard-sphere crystals. Phys Rev Lett 2003;90:255501.
[10] Wang F, Wang ZR, Peng Y, Zheng ZY, Han YL. Homogeneous melting near the superheat limit of hard-sphere crystals. Soft Matter 2018;14:2447.
[11] Lemarchand CA. Molecular dynamics simulations of a hard sphere crystal and reaction-like mechanism for homogeneous melting. J Chem Phys 2012;136:234505.
[12] Hopkins AB, Jiao Y, Stillinger FH, Torquato S. Phase diagram and structural diversity of the densest binary sphere packings. Phys Rev Lett 2011;107:125501.
[13] Torquato S, Jiao Y. Effect of dimensionality on the continuum percolation of overlapping hyperspheres and hypercubes. II. Simulation results and analyses. J Chem Phys 2012;137:074106.
[14] Xu WX, Zhang KX, Zhang YF, Jiang JY. Packing fraction, tortuosity and permeability of granular-porous media with densely packed spheroidal particles: monodisperse and polydisperse systems. Water Resour Res 2022;58:e2021WR031433.
[15] Berthier L. Self-induced heterogeneity in deeply supercooled liquids. Phys Rev Lett 2021;127:088002.
[16] Torquato S. Hyperuniform states of matter. Phys Rep 2018;745:1.
[17] Yang L, Bergstrom ZJ, Wirth BD. Effect of interatomic potential on the energetics of hydrogen and helium-vacancy complexes in bulk, or near surfaces of tungsten. J Nucl Mater 2018;512:357.
[18] Wen M. A new interatomic potential describing Fe-H and H-H interactions in bcc iron. Comput Mater Sci 2021;197:110640.
[19] Daw MS, Baskes MI. Semiempirical, quantum mechanical calculation of hydrogen embrittlement in metals. Phys Rev Lett 1983;50:1285.
[20] Myers SM, et al. Hydrogen interactions with defects in crystalline solids. Rev Modern Phys 1992;64:559.
[21] Reith D, Pütz M, Müller-Plathe F. Deriving effective mesoscale potentials from atomistic simulations. J Comput Chem 2003;24:1624.
[22] Sherman ZM, Howard MP, Lindquist BA, Jadrich RB, Truskett TM. Inverse methods for design of soft materials. J Chem Phys 2020;152:140902.
[23] Chandler D. Introduction To modern statistical mechanics. New York: Oxford University Press; 1987.
[24] Frenkel D, Smith B. Understanding molecular simulation from algorithms to applications. 2nd ed.. New York: Academic Press; 2002.
[25] Ashton DJ, Wilding NB. Three-body interactions in complex fluids: Virial coefficients from simulation finite-size effects. J Chem Phys 2014;140:244118.
[26] Henderson RL. A uniqueness theorem for fluid pair correlation functions. Phys Lett A 1974;49:197.
[27] Lyubartsev AP, Laaksonen A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. Phys Rev E 1995;52:3730.
[28] Torquato S, Wang H. Precise determination of pair interactions from pair statistics of many-body systems in and out of equilibrium. Phys Rev E 2022;106:044122.
[29] Stones AE, Dullens RPA, Aarts DGAL. Model-free measurement of the pair potential in colloidal fluids using optical microscopy. Phys Rev Lett 2019;123:098002.
[30] Lennard-Jones JE. On the determination of molecular fields. —II. From the equation of state of a gas. Proc Poy Soc A 1924;106:463.
[31] Mie G. To the kinetics theory of one-atomic corps. Ann Physics 1903;11:657.
[32] Hansen JP, McDonald IR. Theory of simple liquids. 3rd ed.. New York: Academic Press; 2005.
[33] Tian JX, Gui YX. Liquid-gas phase transition to first order of an argon-like fluid modeled by mie potential. Internat J Modern Phys B 2005;19:3161.
[34] Liu C-H, Tao Y, Du Q, Billinge S. Using a machine learning approach to determine the space group of a structure from the atomic pair distribution function. Acta Crystallogr Sect A 2019;75:633.
[35] Egami T, Billinge SJL. Underneath the bragg peaks: structural analysis of complex materials. Oxford: Elsevier Ltd; 2003.
[36] Borgis D, Assaraf R, Rotenberg B, Vuilleumier R. Computation of pair distribution functions and three-dimensional densities with a reduced variance principle. Mol Phys 2013;111:3486.
[37] Rotenberg B. Use the force! reduced variance estimators for densities, radial distribution functions, and local mobilities in molecular simulations. J Chem Phys 2020;153:150902.
[38] Livingstone D, Manallack D, Tetko I. Data modelling with neural networks: Advantages and limitations. J Comput-Aided Mol Des 1997;11:135.
[39] Bourquin J, Schmidli H, van Hoogevest P, Leuenberger H. Advantages of artificial neural networks (ANNs) as alternative modelling technique for data sets showing non-linear relationships using data from a galenical study on a solid dosage form. Eur J Pharm Sci 1998;7:5.
[40] Mulero A, Pierantozzi M, Cachadina I, Nicola G. An artificial neural network for the surface tension of alcohols. Fluid Phase Equilib 2017;449:28.
[41] Mulero A, Cachadina I, Valderrama JO. Artificial neural network for the correlation and prediction of surface tension of refrigerants. Fluid Phase Equilib 2017;451:60.
[42] Allen MP, Tildesley DJ. Computer simulation of liquids. New York: Oxford University Press; 2017.
[43] Agrawal V, Peralta P, Li YY, Oswald J, Agrawal V, Peralta P, et al. A pressure-transferable coarse-grained potential for modeling the shock hugoniot of polyethylene. J Chem Phys 2016;145:104903.
[44] Liu MH, Oswald J. Coarse–grained molecular modeling of the microphase structure of polyurea elastomer. Polymer 2019;176:1.