Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

On the determination of Lennard-Jones parameters for polyatomic molecules†

Huangrui Mo,[abc] Xiaoqing You,*[ab], Kai Hong Luo,[d] and Struan H Robertson[e]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Characterizing the key length and energy scales of intermolecular interactions, Lennard-Jones parameters, i.e., collision diameter and well depth, are prerequisites for predicting transport properties and rate constants of chemical species in dilute gases. Due to anisotropy in molecular structures, Lennard-Jones parameters of many polyatomic molecules are only empirically estimated or even undetermined. This study focuses on determining the effective Lennard-Jones parameters between a polyatomic molecule and a bath gas molecule from interatomic interactions. An iterative search algorithm is developed to find orientation-dependent collision diameters and well depths on intermolecular potential energy surfaces. An orientation-averaging rule based on characteristic variables is proposed to derive the effective parameters. Cross-interaction parameters for twelve hydrocarbons with varying molecular shapes, including long-chain and planar ones, interacting with four bath gases He, Ar, N_2, and O_2 are predicted and reported. Three-dimensional parametric surfaces are constructed to quantitatively depict molecular anisotropy. Algorithmic complexity analysis and numerical experiments demonstrate that the iterative search algorithm is robust and efficient. By using the latest experimental diffusion data, it is found that the proposed orientation-averaging rule improves the prediction of cross-interaction Lennard-Jones parameters for polyatomic molecules, including for long-chain molecules that challenge the consistency of previous methods. By introducing characteristic variables, the present study shows a new route to determining effective Lennard-Jones parameters for polyatomic molecules.

1 Introduction

As the prerequisites for predicting many physical properties of chemical species,[1–3] Lennard-Jones parameters play an important role in simulating physical and chemical phenomena spanning gaseous, liquid, and solid phases.[4–6] Characterizing the key length and energy scales of non-bonding pair interactions, Lennard-Jones parameters are also an important component for many well-known and widely used potentials/force fields such as the Lennard-Jones potentials,[7] MolMod,[8] TraPPE,[9] OPLS-AA,[10] COMPASS[11] and many others. Therefore, their accurate determination is of fundamental importance in chemical kinetics, transport phenomenon, and molecular dynamics simulations.[4,6,12,13]

For instance, high-fidelity modelling of chemically reacting flows of gases requires chemical kinetic, thermodynamic, and transport properties of chemical species and their mixtures over a wide range of temperatures.[6,14,15] While chemical kinetic parameters such as rate constants are needed for describing chemical reaction processes, transport properties such as diffusivity, viscosity, and thermal conductivity are needed for describing mass, momentum, and energy transfer behaviours. To predict rate constants by solving master equations[16–19] and transport properties of chemical species in dilute gas by using the gas kinetic theory, particularly the Chapman-Enskog theory and its higher order corrections,[2,20] the Lennard-Jones parameters of the involved species are essential information.[6,21,22] Although considerable research has been directed to their measurement and theory, Lennard-Jones parameters of many species still bear large uncertainties or remain undetermined.[12,23–28] Meanwhile, sensitivity studies by various researchers[29–31] have revealed that transport properties determined from Lennard-Jones parameters can have large impacts on simulation results such as concentration distributions, flame propagation and extinction, and pollutant production.

The Lennard-Jones parameters of a pair of interacting molecules consist of two components related to intermolecular interactions: the collision diameter, the separation distance at

[a] Center for Combustion Energy, Tsinghua University, Beijing 100084, China; E-mail: xiaoqing.you@tsinghua.edu.cn
[b] Key Laboratory for Thermal Science and Power Engineering of the Ministry of Education, Tsinghua University, Beijing 100084, China
[c] State Key Laboratory of High Temperature Gas Dynamics, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100190, China
[d] Department of Mechanical Engineering, University College London, Torrington Place, London WC1E 7JE, U.K.
[e] Dassault Systèmes, BIOVIA, 334, Cambridge Science Park, Cambridge CB4 0WN, U.K.
† Electronic Supplementary Information (ESI) available: [see attached]

which the intermolecular potential energy equals zero, and the well depth, the minimal value of the intermolecular potential energy. In the literature, optimizing initially guessed potential parameters by using macroscopic thermophysical property data as optimization constraints[4,5] or by using transport data combined with the inversion of the Chapman-Enskog relations[25,32,33] as well as experimental measurements using neutron/x-ray diffraction,[34,35] molecular beam scattering,[26,36] or spectroscopy[37–39] were the primary means to obtain Lennard-Jones parameters. Lennard-Jones parameters optimized from macroscopic data are a posteriori estimates and often involve the empirical Lorentz-Berthelot or modified Lorentz-Berthelot combining rule to describe cross interactions. Meanwhile, parameters obtained from experiments are usually scarce and limited in thermodynamic conditions and molecular types.[12,40] With advances in computational chemistry, it becomes increasingly feasible to directly calculate Lennard-Jones parameters from interatomic interactions described by force fields obtained from quantum mechanical calculations, such as ab initio fitted potentials.[41–43] This first-principle calculation approach can achieve a priori predictions without the need of pre-acquired experimental data and theoretically can be applied to any chemical species.

To determine the self- or cross-interaction Lennard-Jones parameters for polyatomic molecules through first-principle calculations, a major challenge arises from the presence of complex molecular interactions and structural anisotropy.[2,20] Many researchers have attempted to tackle this challenge. Hirschfelder *et al.*[2] described a spherical average method that spherically averages the anisotropic intermolecular interactions to obtain an effective isotropic potential, from which the effective Lennard-Jones parameters can be readily derived. Monchick and Green[44] examined the spherical average method and found good prediction on transport and scattering cross sections for systems like HCl-He. In studying collisional energy transfer in toluene-Ar systems, Lim[45] suggested finding the effective Lennard-Jones parameters for arbitrary intermolecular potentials by averaging the collision diameter and well depth on each relative orientation. Using the benzene-Ar system, Bernshtein and Oref[41] proposed obtaining effective dynamic potential, collision diameter, and well depth by averaging the intermolecular interactions resulting from a large set of collision trajectories. Employing interatomic Buckingham potentials fitted from high-level quantum chemistry calculations, Jasper and Miller[43] evaluated the spherical average, collision diameter and well depth average, and trajectory average approaches and suggested that the collision diameter and well depth average approach, which is referred to as one-dimensional minimization method,[43,46] balances accuracy and efficiency by the explicit treatment of anisotropy while avoiding expensive trajectory calculations. A review of the efforts taken by some of the aforementioned researchers was provided by Lendvay.[46]

Furthermore, the self- and cross-interaction Lennard-Jones parameters for polyatomic molecules are also difficult to evaluate directly. Binary diffusion coefficients can be an ideal reference for evaluating cross-interaction parameters, since the binary diffusion coefficients of two mixing species can be analytically predicted by their effective cross-interaction Lennard-Jones parameters via the Chapman-Enskog theory. In a span of about a decade, McGivern, Manion and co-workers measured the binary diffusion coefficients of normal alkanes ranging from C_1 to C_8 interacting with bath gases He and N_2 using reversed-flow chromatography.[47–50] Employing this consistent and contemporary set of experimental diffusion data,[49,51,52] Liu *et al.*[49] found that predictions by effective Lennard-Jones parameters obtained via averaging collision diameters and well depths on relative orientations show marked deviations from experimental data for long-chain alkanes like n-heptane. To address this limitation, Liu *et al.*[49] proposed an alternative theoretical model for binary diffusion coefficients of long-chain molecules using gas-kinetic theory analysis for slender bodies, which, however, cannot be applied to polyatomic molecules with more general structures. Therefore, resolving the current inconsistencies in the effective Lennard-Jones parameters between polyatomic molecules and bath gas molecules is of theoretical interest and practical importance.

When a polyatomic molecule interacts with another molecule, the intermolecular interaction and related potential parameters on each relative orientation can be considered as certain and inherently determined by electronic structures. To better capture the effective Lennard-Jones parameters between a polyatomic molecule and a bath gas molecule, the underlying task could lie in finding a proper averaging rule to account for the statistical effects of molecular rotations and randomly oriented collisions during a large number of molecular encounters. The present work aims to explore the existence of such averaging rules. By introducing two characteristic variables representing repulsive and attractive energy scales, a new orientation-averaging rule based on the characteristic variables is proposed. By using the complete set of experimental binary diffusion data of McGivern, Manion and co-workers,[47–50] it will be shown that this new rule improves the prediction of effective Lennard-Jones parameters for polyatomic molecules interacting with bath gases and resolves the inconsistency of the Chapman-Enskog theory based diffusion prediction for long-chain molecules. Meanwhile, an iterative search algorithm is developed to robustly and efficiently find orientation-dependent collision diameters and well depths on complex intermolecular potential energy surfaces, which provides prerequisite data for the averaging step. In addition, three-dimensional parametric surfaces of Lennard-Jones parameters are constructed in this study to quantitatively depict anisotropic interactions in polyatomic molecules.

2 Theory and methods

Lennard-Jones parameters, namely collision diameter σ and well depth ε, are the basic parameters of the Lennard-Jones m-6 potentials, which describe orientation-independent intermolecular interaction as the following form

$$\phi(r) = \varepsilon \left(\frac{m}{m-6}\right)\left(\frac{m}{6}\right)^{6/(m-6)}\left[\left(\frac{\sigma}{r}\right)^m - \left(\frac{\sigma}{r}\right)^6\right] \qquad (1)$$

where $\phi(r)$ denotes the intermolecular potential energy, r is the separation distance between the centers of mass of the molecules, σ is the collision diameter at which $\phi(r=\sigma)=0$, and $\varepsilon = -\min\{\phi(r)\} = -\phi(r=r_{\min})$ is the depth of the potential well at r

$= r_{\min}$, m is an integer such as $m = 9$, 12, … and represents the steepness of the repulsive energy.

As σ and ε represent the root and minimum of intermolecular potential energy respectively, they can also be linked to other widely used potentials such as the Buckingham potentials[53,54]

$$\phi(r) = \frac{\varepsilon}{1-\frac{6}{\alpha}}\left\{\frac{6}{\alpha}\exp\left[\alpha\left(1-\frac{r}{r_{\min}}\right)\right]-\left(\frac{r_{\min}}{r}\right)^6\right\} \tag{2}$$

where ε and $r_{\min}$ remain the well depth and well depth location respectively, $r_{\min}$ can be related to σ by $\phi(r=\sigma)=0$, such that

$$6\ln\frac{\sigma}{r_{\min}} - \alpha\frac{\sigma}{r_{\min}} = \ln\frac{\alpha}{6} - \alpha \tag{3}$$

and α is an additional parameter used to measure the steepness of the repulsive energy.

Characterizing the length and energy scales of the intermolecular interaction of a molecular pair, Lennard-Jones parameters also play an integral role in the derivation of many fundamental kinetic and transport quantities, such as hard sphere collision cross section $\pi\sigma^2$, reduced temperature $T^* = k_B T/\varepsilon$, collision integrals $\Omega^{(l,s)}$ ($l = 1,2,\ldots$, $s = 1,2,\ldots$), and total collision rate $\sqrt{8k_BT/(\pi\mu)}\pi\sigma^2\Omega^{(2,2)}$,[1,2,15,55] where k_B is the Boltzmann constant and μ is reduced mass.

One first-principle approach for obtaining Lennard-Jones parameters is to calculate the parameters from ab initio interatomic potentials.[41,43] However, as illustrated in Fig. 1, for an interacting molecular pair A + M, where A and M denote the target species and bath gas respectively, their intermolecular potential is a function of relative position, orientation, and internal degrees of freedom.[38,56] Consequently, the determination of the corresponding intermolecular σ and ε is not straightforward and requires dedicated treatment of the anisotropic effects in intermolecular potential energy surface.

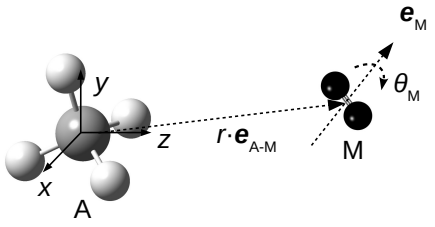


Fig. 1 Schematic diagram illustrating a target molecule A interacting with a bath gas molecule M.

Since intermolecular interactions are generally much weaker than intramolecular interactions,[1,56] a separable approximation can be adopted to describe the total potential energy surface of A + M:

$$\phi_{A+M}(R_A, R_M, r, \Omega) = \phi_A(R_A) + \phi_M(R_M) + \phi_{A-M}(r,\Omega) \tag{4}$$

where ϕ_{A+M} is the total potential energy of A + M, R_A and R_M represent the internal coordinates of the target species A and bath gas M respectively, r and Ω denote the separation between centers of mass and relative orientation of A and M respectively, ϕ_A and ϕ_M are the intramolecular potential of isolated A and M respectively, ϕ_{A-M} is the intermolecular potential between A and M and is reduced to a function of r and Ω, i.e., $\phi_{A-M} = \phi_{A-M}(r,\Omega)$.

To describe the anisotropic intermolecular potential ϕ_{A-M}, a sum of pairwise atom-atom interactions can be used to approximate the potential energy surface:

$$\phi_{A-M}(r,\Omega) = \sum_{i\in A}\sum_{j\in M}\phi_{ij}(r_{ij}) \tag{5}$$

and the interatomic potential $\phi_{ij}(r_{ij})$ can be adequately described by a modified Buckingham potential,[43] referred to as exp-6,

$$\phi_{ij}(r_{ij}) = E_{ij}\exp(-r_{ij}/B_{ij}) - C_{ij}^6/(r_{ij}^6 + D_{ij}^6) \tag{6}$$

where E_{ij}, B_{ij}, C_{ij}, and D_{ij} are fitted parameters for interatomic interactions. ij represents the interacting binary atoms with one atom from the target molecule A and the other from the bath gas molecule M, and r_{ij} is the interatomic distance.

In order to derive the effective collision diameter σ and well depth ε for the intermolecular potential ϕ_{A-M}, ensembles of A + M pairs are needed to sample the anisotropic intermolecular potential energy surface, which lead to a set of orientation-dependent values of $\phi_{A-M}(r,\Omega)$. The idea of the spherical average method[44] is to average the potentials over orientations to obtain an isotropic potential:

$$\phi_{SA}(r) = \sum_{n=1}^{N_\Omega}\phi_{A-M}(r,\Omega_n)/N_\Omega \tag{7}$$

where Ω_n denotes the n-th sampled orientation, and N_Ω is the total number of sampled orientations. The effective collision diameter and well depth can then be derived as

$$\phi_{SA}(\sigma) = 0 \tag{8a}$$

$$\varepsilon = -\min\{\phi_{SA}(r)\} \tag{8b}$$

Another widely used method is to first determine the collision diameter σ_n and well depth ε_n for $\phi_{A-M}(r,\Omega_n)$ with orientation Ω_n, $n = 1,\ldots,N_\Omega$, and then average the collision diameters and well depths over orientations to obtain the effective collision diameter and well depth:[43,45]

$$\sigma = \sum_{n=1}^{N_\Omega}\sigma_n/N_\Omega \tag{9a}$$

$$\varepsilon = \sum_{n=1}^{N_\Omega}\varepsilon_n/N_\Omega \tag{9b}$$

In determining σ_n and ε_n, the simplest solution is to directly scan the potential curve $\phi_{A-M}(r,\Omega_n)$ using N_s uniformly distributed sample points. The computational load then is $W_{scn} = O(N_\Omega N_s)$, where N_Ω is the number of orientations, and N_s is the number of sample points per orientation. When N_s is increased, to obtain a higher accuracy for σ_n and ε_N, W_{scn} will increase substantially due to the scaling factor by N_Ω. Therefore, it is necessary to develop iterative search methods for finding σ_n and ε_n on Ω_n

with much higher efficiency.

As shown in Fig. 2, when a polyatomic molecule A, such as $C_{24}H_{12}$, interacts with a bath gas molecule M, such as He, the intermolecular potential $\phi_{\mathrm{A-M}}(r,\Omega_n)$ built on interatomic potentials, such as exp-6, can possess the following numerical characteristics:

$$\phi_{\mathrm{A-M}}(r,\Omega_n) \to 0,\ r \geq r_c \tag{10a}$$

$$\phi_{\mathrm{A-M}}(r,\Omega_n) < 0,\ \sigma_n < r < r_c \tag{10b}$$

$$\phi_{\mathrm{A-M}}(r,\Omega_n) = 0,\ r = \sigma_n \tag{10c}$$

$$\phi_{\mathrm{A-M}}(r,\Omega_n) > 0,\ r_m < r < \sigma_n \tag{10d}$$

$$\phi_{\mathrm{A-M}}(r,\Omega_n) \text{ oscillates},\ 0 < r \leq r_m \tag{10e}$$

and the intermolecular force $f = -\mathrm{d}\phi/\mathrm{d}r$ correspondingly has the following numerical characteristics:

$$f_{\mathrm{A-M}}(r,\Omega_n) \to 0,\ r \geq r_c \tag{11a}$$

$$f_{\mathrm{A-M}}(r,\Omega_n) < 0,\ \ r_{\min} < r < r_c \tag{11b}$$

$$f_{\mathrm{A-M}}(r,\Omega_n) = 0,\ r = r_{\min} \tag{11c}$$

$$f_{\mathrm{A-M}}(r,\Omega_n) > 0,\ r_m < r < r_{\min} \tag{11d}$$

$$f_{\mathrm{A-M}}(r,\Omega_n) \text{ oscillates},\ 0 < r \leq r_m \tag{11e}$$

in which r_c represents the value of intermolecular separation at which $\phi_{\mathrm{A-M}}(r,\Omega_n)$ and $f_{\mathrm{A-M}}(r,\Omega_n)$ are smaller than a given convergence tolerance for root finding, and r_m represents the value of intermolecular separation at which $\phi_{\mathrm{A-M}}(r,\Omega_n)$ and $f_{\mathrm{A-M}}(r,\Omega_n)$ falsely become non-positive. The existence of r_m can be caused by the overlap of atoms between A and M at a small r or by the Buckingham catastrophe[2] in the exp-6 interatomic potential, which causes the intermolecular potential and force to approach $-\infty$ for $r \to 0$.

In the intermolecular potential energy surface, the existence of unphysical r_m and r_c can severely undermine the robustness of an iterative search method for σ_n and ε_n, which are the roots of the intermolecular potential and force, respectively. Jasper and Miller[57] explored a Newton-Raphson technique to improve search efficiency but noted that the solution process would break for short molecular distances. Considering that initial guesses or intermediate calculations in the region $0 < r \leq r_m$ or $r \geq r_c$ can easily break a root search method, and r_m and r_c cannot be known a priori and change with Ω_n, in order to develop a robust method, a projection preprocessing is proposed in this study to determine the lower bound of the initial iteration interval $[r_l, r_u]$ such that

$$r_l = \max\{r_{\mathrm{A}-i}\cdot e_{\mathrm{A-M}}\} + \Delta r + \max\{r_{\mathrm{M}-j}\cdot e_{\mathrm{M-A}}\},\ \text{for } \forall i \in \mathrm{A},\ \forall j \in \mathrm{M} \tag{12}$$

where $r_{\mathrm{A}-i} = r_i - r_{\mathrm{A}}$ is the position vector of atom i in A relative to the center of mass of A; similarly, $r_{\mathrm{M}-j} = r_j - r_{\mathrm{M}}$ is the position vector of atom j in M relative to the center of mass of M, $e_{\mathrm{A-M}} = (r_{\mathrm{M}} - r_{\mathrm{A}})/|r_{\mathrm{M}} - r_{\mathrm{A}}|$ is the unit direction vector from the center of mass of A to that of M, $e_{\mathrm{M-A}} = -e_{\mathrm{A-M}}$, and $\Delta r = 0.1\,\text{Å}$ is a small distance used to avoid molecular overlap.

The determined lower bound r_l guarantees that the closest two atoms from A and M for any orientation Ω_n will have a minimum distance of Δr, at which the intermolecular potential and force will generally be positive. However, practical tests on a set of A + M revealed that, when A is a highly planar molecule such as coronene, about 5% of sampled orientations still showed negative values. Therefore, an extra local search step is used to further adjust the value of r_l, in which the value of $\phi_{\mathrm{A-M}}(r_l,\Omega_n)$ (when solving for σ_n) or $f_{\mathrm{A-M}}(r_l,\Omega_n)$ (when solving for ε_n) is first checked, if it is positive, then r_l is well determined; otherwise, search in the interval $r \in [r_l, r_l + 5]$ along the direction $e_{\mathrm{M-A}}$ with step size $2\Delta r$ until $\phi_{\mathrm{A-M}}(r,\Omega_n)$ or $f_{\mathrm{A-M}}(r,\Omega_n)$ is positive, and r_l is then determined as the value of r corresponding to the termination step. Typically, this extra local search process takes about 10 steps for those initially non-positive $\phi_{\mathrm{A-M}}(r_l,\Omega_n)$ or $f_{\mathrm{A-M}}(r_l,\Omega_n)$.

For the upper bound r_u of the initial iteration interval $[r_l, r_u]$, as $\phi_{\mathrm{A-M}}(r,\Omega_n)$ or $f_{\mathrm{A-M}}(r,\Omega_n)$ usually features a close-to-zero long tail for relatively large r, the search for $|\phi_{\mathrm{A-M}}(r_u,\Omega_n)| > e_c$ or $|f_{\mathrm{A-M}}(r_u,\Omega_n)| > e_c$, where e_c is the chosen convergence tolerance, can be expensive and dependent on the value of e_c. Therefore, no special requirement on r_u is imposed except that r_u is large enough, for example, $r_u = 20\,\text{Å}$, such that σ_n and ε_n are initially included in the iteration interval $[r_l, r_u]$. Consequently, a large portion of the target function $|\phi_{\mathrm{A-M}}(r,\Omega_n)|$ or $|f_{\mathrm{A-M}}(r,\Omega_n)|$ in the initial iteration interval $[r_l, r_u]$ can be well within the convergence tolerance or even be zero. In order to ensure the correct convergence to σ_n or ε_n rather than to a random point in the long tail region, a modified bisection search is proposed for root searching over the iteration interval $[r_l, r_u]$:

Step 1 Calculate $N_s = \log_2[(r_u - r_l)/e_r]$, where e_r is a predefined error tolerance for root searching;

Step 2 Compute $r_k = (r_l + r_u)/2$;

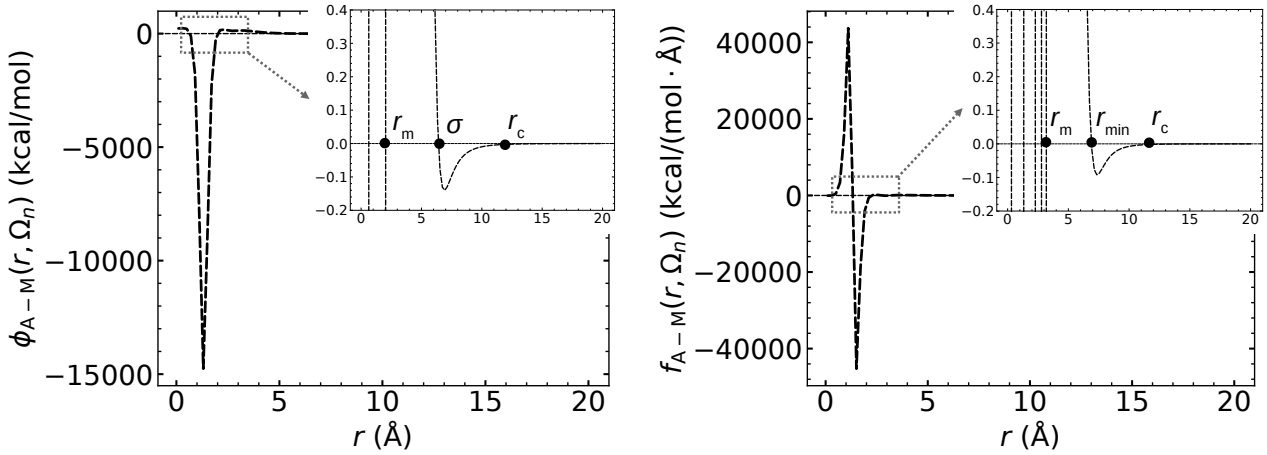
Step 3 If $\phi_{\mathrm{A-M}}(r_k,\Omega_n)\cdot\phi_{\mathrm{A-M}}(r_l,\Omega_n) > 0$ (when solving for σ_n) or $f_{\mathrm{A-M}}(r_k,\Omega_n)\cdot f_{\mathrm{A-M}}(r_l,\Omega_n) > 0$ (when solving for ε_n), set $r_l = r_k$ and keep r_u, otherwise, set $r_u = r_k$ and keep r_l;

Step 4 $k = k + 1$;

Step 5 Repeat steps 2-4 until $k \geq N_s$.

This modified bisection search will guarantee convergence to σ_n or ε_n within the error tolerance of e_r, even when the target function has many close-to-zero/zero values in the long tail region, since the length of the iteration interval will be $(r_u - r_l)/2^{N_s} \leq e_r$ after N_s iterations and the Step 3 ensures that σ_n or ε_n is always within the iteration interval.

After successfully finding the collision diameters and well depths of individual orientations, which can be considered as certain and inherently determined by electronic structures, the next critical step is to determine the effective values. Concerning the inconsistencies in using the Chapman-Enskog theory with effective Lennard-Jones parameters obtained by directly averaging the orientation-dependent σ_n and ε_n, it is of theoretical interest and practical importance to explore whether alternative



(a) (b)

Fig. 2 Calculated intermolecular interactions for $C_{24}H_{12}$ + He pair at one random orientation Ω_n illustrating the unphysical behaviours caused by interatomic potentials and overlapping atoms. (a) Intermolecular potential curve. (b) Intermolecular force curve. The core regions are shown at the top right corners.

averaging rules can better resolve anisotropic effects for polyatomic molecules. In this study, new characteristic variables $\eta = \varepsilon\sigma^{12}$ and $\xi = \varepsilon\sigma^{6}$ as well as the Boltzmann weighting factor $w = \exp(\varepsilon/k_B T)$ with $T = 298\,\mathrm{K}$ are introduced, in which η and ξ represent the strength of the repulsive and dispersive interactions in the Lennard-Jones 12-6 potential form respectively. In addition to the conventional spherical average in Eq. (7) (denoted as SA) and σ-ε-based arithmetic average in Eq. (9) (denoted as σ-ε-AM), three additional average rules based on the introduced variables and weight are constructed, which are η-ξ-based arithmetic average (denoted as η-ξ-AM):

$$\eta = \sum_{n=1}^{N_\Omega} \eta_n / N_\Omega \tag{13a}$$

$$\xi = \sum_{n=1}^{N_\Omega} \xi_n / N_\Omega \tag{13b}$$

$$\sigma = (\eta/\xi)^{1/6} \tag{13c}$$

$$\varepsilon = (\xi^2/\eta) \tag{13d}$$

σ-ε-based Boltzmann weighting (denoted as σ-ε-BW):

$$\sigma = \sum_{n=1}^{N_\Omega} \sigma_n w_n / \sum_{n=1}^{N_\Omega} w_n \tag{14a}$$

$$\varepsilon = \sum_{n=1}^{N_\Omega} \varepsilon_n w_n / \sum_{n=1}^{N_\Omega} w_n \tag{14b}$$

and η-ξ-based Boltzmann weighting (denoted as η-ξ-BW):

$$\eta = \sum_{n=1}^{N_\Omega} \eta_n w_n / \sum_{n=1}^{N_\Omega} w_n \tag{15a}$$

$$\xi = \sum_{n=1}^{N_\Omega} \xi_n w_n / \sum_{n=1}^{N_\Omega} w_n \tag{15b}$$

$$\sigma = (\eta/\xi)^{1/6} \tag{15c}$$

$$\varepsilon = (\xi^2/\eta) \tag{15d}$$

In this study, to calculate the effective collision diameter and well depth, ensembles of A + M pairs are prepared via a Monte Carlo sampling procedure designed as follows. As illustrated in Fig. 1, the total geometry of A + M depends on the internal coordinates R_A and R_M, the relative orientation Ω, and the center-of-mass separation r. R_A and R_M are obtained from equilibrium structures, which ignores multiconformers and temperature-dependent vibrations of molecules, as their effects on the final ensemble averages are generally small.[43] As a result, the introduced Boltzmann weighting in the σ-ε-BW and η-ξ-BW rules can be regarded as approximating the expensive trajectory average method[41] by assigning larger weights to intermolecular interactions with deeper potential wells. During the Monte Carlo sampling, the center of mass of R_A is held fixed at the origin of the coordinate system with the principal moment of inertia axes aligned with the z-, y-, x-axis in the order of increasing principal moment of inertia. The orientation Ω is equated as $\Omega \equiv (e_{A-M}, e_M, \theta_M)$, in which e_{A-M} is the unit direction vector pointing from the center of mass of A to the center of mass of M and is randomly sampled from a uniform distribution in space, e_M is a randomly sampled unit vector passing the center of mass of M and acts as the self-rotating axis of M, and θ_M is the rotational angle of M about the axis e_M and is randomly sampled from a uniform distribution in $[0, 2\pi]$. At each orientation Ω_n, two r samples at $r = r_l$ and $r = r_u$

are initially generated via the projection preprocessing described before. It is worth noting that the orientation definition Ω defined in Fig. 1 consists of five degrees of freedom (two in e_{A-M}, two in e_M, and one in θ_M) and applies to a system of two nonlinear tops. For other types of systems the number of degrees of freedom represented by Ω is progressively reduced as the complexity of colliders are altered from nonlinear top, to linear top and, finally, to atom. For instance, when A is a nonlinear molecule like CH_4 and M is a monatomic molecule like He, the self-rotating vector e_M and angle θ_M are no longer effective, and so Ω is then uniquely determined by the two degrees of freedom in e_{A-M}.

Effective Lennard-Jones parameters for a set of 48 A + M target-bath pairs are discussed, in which A = methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), ethane (C_2H_6), propane (C_3H_8), *n*-butane (*n*-C_4H_{10}), *n*-pentane (*n*-C_5H_{12}), *n*-hexane (*n*-C_6H_{14}), *n*-heptane (*n*-C_7H_{16}), *n*-octane (*n*-C_8H_{18}), *n*-nonane (*n*-C_9H_{20}), coronene ($C_{24}H_{12}$) and M = He, Ar, N_2, O_2. The equilibrium structures of the molecules are obtained from the NIST Computational Chemistry Comparison and Benchmark Database.[58] The geometric shapes of target molecules can be roughly categorized as quasi-spherical (CH_4), linear (C_2H_2), planar (C_2H_4, $C_{24}H_{12}$), and straight-chain (C_nH_{2n+2}, $n = 2,\ldots,9$). Python was used to perform the ensemble generation, to integrate the Large-scale Atomic/Molecular Massively Parallel Simulator[59] (LAMMPS) for evaluating the intermolecular potential $\phi_{A-M}(r,\Omega)$, and to implement the aforementioned five averaging rules for obtaining effective Lennard-Jones parameters. Since only intermolecular interactions need to be sampled from static molecular configurations, the intramolecular potentials are not required. The interatomic potential parameters E_{ij}, B_{ij}, C_{ij}, and D_{ij} in Eq. (6) are obtained from Ref.[60] and are tabulated in the Supporting Information. These interatomic parameters were fitted from MP2 and QCISD(T) level of energy calculations based on CH_4 + M systems and assumed to be transferable to other hydrocarbon targets.[43] For each A + M system, $N_\Omega = 10{,}000$ random orientations $\{\Omega_n\}$ were sampled and evaluated. For the systems considered in this study, the one-sigma uncertainties of the effective collision diameters and well depths estimated using the bootstrap method[61] are converged within 0.3% for the chosen N_Ω, even for the most anisotropic ones.

3 Results and discussion

3.1 Efficiency comparison and analysis

When finding collision diameters and well depths for A + M on interval $[r_l, r_u]$ with number of orientations N_Ω and spatial error e_r, the computational load for a direct scanning approach is $W_{scn} = O(N_\Omega(r_u - r_l)/e_r)$, while the proposed iterative search method has $W_{itr} = O(N_\Omega \log_2[(r_u - r_l)/e_r])$, where, $O(n)$ represents the evaluation of the intermolecular potential for n times. The change of algorithmic complexity from W_{scn} to W_{itr} can result in a significant reduction of computational cost.

To further demonstrate the efficiency benefits, numerical experiments were conducted for the $C_{24}H_{12}$ + He pair on the initial sampling interval $[r_l = 0.1\text{Å}, r_u = 20\text{Å}]$. The analytical computation load and practical computation time of the direct scanning and iterative search method are compared and shown in Table 1, in which N_Ω is the predefined number of orientations, e_r is the predefined spatial error for position sampling on each orientation, W_{scn} and W_{itr} are the analytically derived computational loads described above, and t_{scn} and t_{itr} are the practical computation time for the direct scanning and iterative search method, respectively.

Results show that, for a given N_Ω, an order of magnitude of decrease of e_r can lead to an order of magnitude of increase of the required computation time for the direct scanning approach, while the corresponding increase of computation time is only about 1.2 times for the iterative search method. With the same N_Ω and e_r, the ratio of the practical computation time t_{itr} / t_{scn} closely agrees with the ratio of analytical computation load W_{itr}/W_{scn} for all the tested N_Ω and e_r, in which the small discrepancies are mainly caused by the extra preprocessing steps. While the determined effective parameters are identical within the resolution range of e_r, the required t_{itr} can be orders of magnitude smaller than t_{scn}, particularly for a small e_r. For instance, when $N_\Omega = 10^4$ and $e_r = 0.2\text{Å}$, t_{itr} is about 9.6% of t_{scn}, and when $N_\Omega = 10^2$ and $e_r = 0.002\text{Å}$, t_{itr} is only about 0.15% of t_{scn}. The efficiency gain of the iterative search method can be further enlarged with decreasing e_r, which represents increasing spatial accuracy of the sampled parameters. When using $e_r = 0.2\text{Å}$, which represents 100 samples per orientation for the direct scanning approach, if the obtained solutions of σ_n are substituted into Eq. (6) for evaluating the residual intermolecular potential, it is found that the residual intermolecular potentials at some orientations can be evaluated at about 100 K, which are relative far from the supposed zero potential value. For $e_r = 0.002\text{Å}$ and $N_\Omega = 10^4$, the computational load $W_{scn} = O(10^8)$ would become too expensive for the direct scanning approach, but $W_{itr} = O(1.4\times10^5)$ is still very manageable for the iterative search method. These results demonstrate the efficiency and robustness of the proposed iterative search method for finding orientation-dependent collision diameters and well depths.

3.2 Effective collision diameters and well depths

The calculated effective collision diameters and well depths of A + M systems by different averaging rules are shown in Table 2. The empirical Lennard-Jones parameters for A + M in Table 2 were obtained according to the following procedure: For a self-interaction A + A system, the combined viscosity and second virial coefficient fitted parameters from Tee *et al.*[32] were adopted if available. Otherwise, the viscosity fitted parameters from Hirschfelder *et al.*[2] and the correlation estimated parameters from Wang and Frenklach[25] were used accordingly. The currently chosen empirical parameters are classical and widely used in transport property calculation packages such as TRANLIB.[62] The empirical cross-interaction parameters for A + M were then derived by the Lorentz-Berthelot combining rules[12] using the A + A and M + M parameters

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \tag{16}$$

Table 1 Comparison of analytical computation load and practical computation time of the direct scanning and iterative search method for finding orientation-dependent collision diameters and well depths

N_Ω	e_r	W_{itr}	W_{scn}	$W_{\text{itr}}/W_{\text{scn}}$	$t_{\text{itr}}(\text{s})$	$t_{\text{scn}}(\text{s})$	t_{itr} / t_{scn}
10^2	2.0×10^{-1}	$\text{O}(7.0\times10^2)$	$\text{O}(10^4)$	$\mathbf{7.0\times10^{-2}}$	8.1×10^1	7.6×10^2	$\mathbf{1.1\times10^{-1}}$
10^3	2.0×10^{-1}	$\text{O}(7.0\times10^3)$	$\text{O}(10^5)$	$\mathbf{7.0\times10^{-2}}$	7.8×10^2	7.6×10^3	$\mathbf{1.0\times10^{-1}}$
10^4	2.0×10^{-1}	$\text{O}(7.0\times10^4)$	$\text{O}(10^6)$	$\mathbf{7.0\times10^{-2}}$	7.8×10^3	8.1×10^4	$\mathbf{9.6\times10^{-2}}$
10^2	2.0×10^{-2}	$\text{O}(1.0\times10^3)$	$\text{O}(10^5)$	$\mathbf{1.0\times10^{-2}}$	9.7×10^1	7.4×10^3	$\mathbf{1.3\times10^{-2}}$
10^3	2.0×10^{-2}	$\text{O}(1.0\times10^4)$	$\text{O}(10^6)$	$\mathbf{1.0\times10^{-2}}$	9.8×10^2	7.7×10^4	$\mathbf{1.3\times10^{-2}}$
10^2	2.0×10^{-3}	$\text{O}(1.4\times10^3)$	$\text{O}(10^6)$	$\mathbf{1.4\times10^{-3}}$	1.2×10^2	7.8×10^4	$\mathbf{1.5\times10^{-3}}$

and

$$\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{\frac{1}{2}} \tag{17}$$

in which ij, ii, and jj denote the parameters for A+M, A+A, and M+M, respectively. The parameters for M+M, M = He, Ar, N_2, O_2, were obtained from Jasper and Miller[43] and are tabulated in the Supporting Information.

A direct comparison between the predicted Lennard-Jones parameters by different averaging rules and the empirical parameters suggests that the agreement on collision diameters is much better than that on well depths. For the majority of normal alkane-bath gas pairs, such as $C_nH_{2n+2}+M$ with $n \leq 7$, the deviations in the predicted collision diameters and well depths by the conventional σ-ε-AM rule are within 5% and 20% of the empirical ones, respectively. However, deviations can vary significantly from the empirical parameters for large molecules, such as about 40% over-prediction in collision diameter and about 80% under-prediction in well depth for $C_nH_{2n+2}+M$ pairs, $n > 7$.

The reason for the above wide range of deviations is complicated. A major issue is that a direct comparison between the Lennard-Jones parameters predicted from atomic interactions and the parameters empirically derived from macroscopic viscosity data may not be very useful in evaluating the accuracy of the predicted parameters, due to difficulties like the non-uniqueness and uncertainties in deriving the empirical Lennard-Jones parameters via Chapman-Enskog relations.[12,40,63] Therefore, a much more robust comparison using the latest experimental diffusion data is described later to evaluate the effectiveness of the predicted effective parameters by different averaging rules as well as the collected empirical parameters.

3.3 Parametric surfaces and anisotropic effects

Relatively little effort has been directed toward quantitatively revealing the anisotropic effects in collision diameters and well depths of complex molecules and their relation to molecular structures. Taking advantage of the developed iterative search algorithm and the designed Monte Carlo sampling procedure in this study, three-dimensional parametric surfaces of collision diameters and well depths can be quantitatively constructed and visualized. These parametric surfaces can characterize the spatial properties of the intermolecular potential energy surfaces and enrich the understanding of anisotropic interactions. Figure 3 presents the three-dimensional parametric surfaces of the collision diameter σ, well location r_{min}, and well depth ε for several target-He pairs. The complete set of 144 parametric surfaces for all the target-bath pairs studied here is given in the Supporting Information.

It is observed that the shapes of collision diameter surfaces (the left column in Fig. 3) closely correlate with the overall shapes of the target molecules, and the well location surfaces (the middle column in Fig. 3) are generally enlarged versions of the collision diameter surfaces. The well depths surfaces (the right column in Fig. 3) appear to be influenced by more factors, such as the overall shapes of molecules, types of atoms, and the structural arrangement of a given atom type, and thus reveal more complicated spatial distributions.

For normal alkanes with either quasi-spherical (such as CH_4) or straight-chain shapes (such as n-C_9H_{20}), the global minima of well depths usually appear at hydrogen atoms that are farthest away from the center of mass, and global maxima of well depths tend to locate between those hydrogen atoms. Compared with normal alkanes, a complexity related to structural arrangement of atoms can be observed for molecules like $C_{24}H_{12}$. When the target molecule is a normal alkane, the bath gas molecule will mainly interact with hydrogen atoms, as the carbon atoms are mostly obscured by hydrogen atoms. When the target molecule is changed from a normal alkane to a planar $C_{24}H_{12}$, a large portion of interactions will be directed to unhindered carbon atoms. Consequently, bath gas interactions with both hydrogen atoms and carbon atoms can impact the well depth surface of $C_{24}H_{12}$, and this structural arrangement effect imposes more requirements in optimizing interatomic potential parameters.

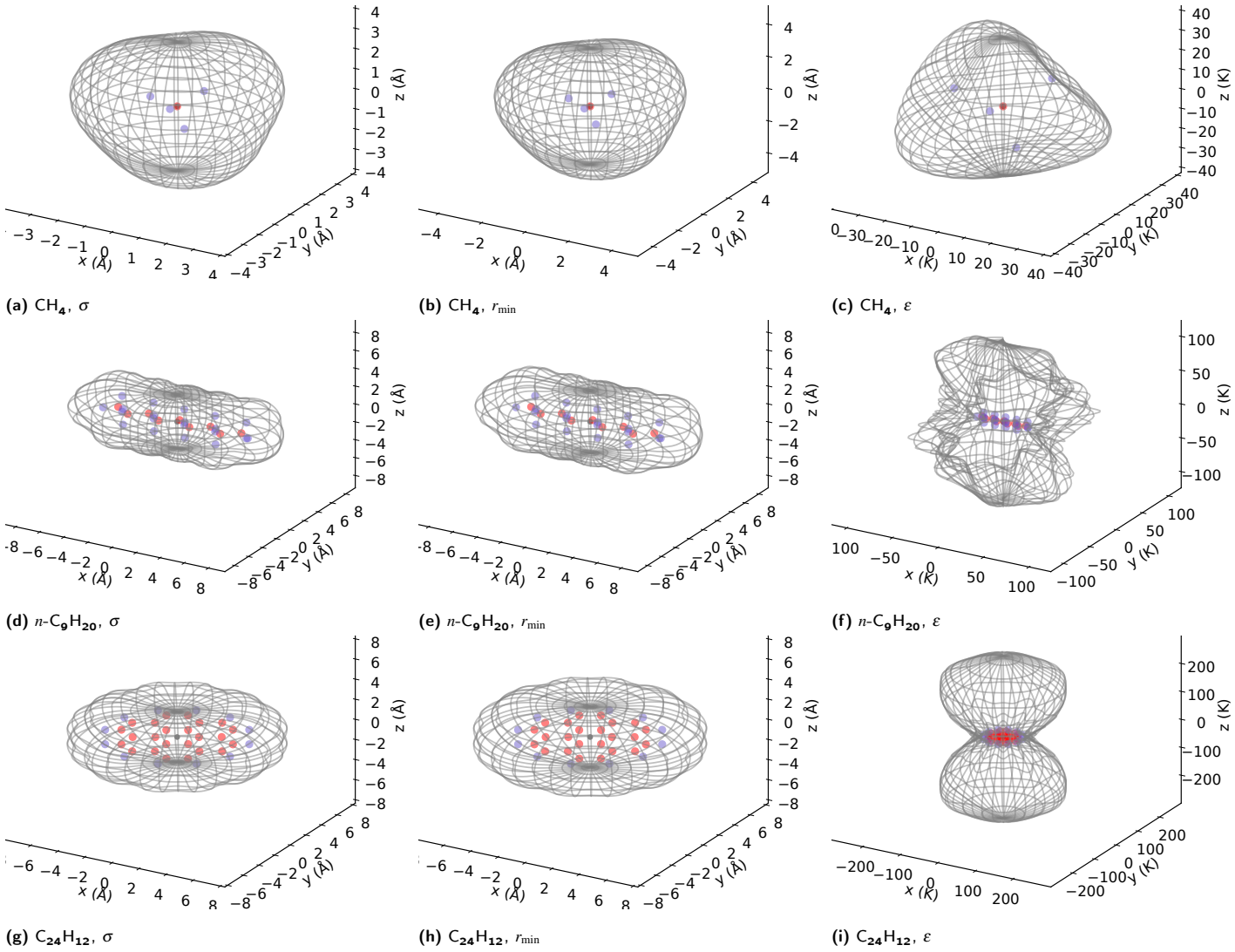
From the constructed quantitative parametric surfaces, it is evident that the size, shape, and atom types of molecules can largely influence the anisotropy of intermolecular interactions and hence the Lennard-Jones parameters. Straight-chain and planar molecules with large sizes tend to have strong anisotropies. Therefore, these molecules present much more challenges and hence can better differentiate the accuracy of calculation methods and averaging rules in determining effective parameters from the orientation-dependent collision diameters and well depths.

3.4 Evaluation on accuracy and averaging rules

To examine the Lennard-Jones parameters in Table 2, an ideal reference can be binary diffusion coefficient, which represents one

Table 2 Calculated effective collision diameters and well depths by different averaging rules with empirical parameters for a set of target-bath pairs

Target	Bath	Averaging rules										Ref. data	
		σ-ε-AM		η-ξ-AM		σ-ε-BW		η-ξ-BW		SA		Empirical	
		σ (Å)	ε (K)	σ (Å)	ε (K)	σ (Å)	ε (K)	σ (Å)	ε (K)	σ (Å)	ε (K)	σ (Å)	ε (K)
CH_4	He	3.32	30.9	3.37	27.9	3.32	31.0	3.37	28.0	3.40	27.7	3.13	41.3[32]
	Ar	3.53	163.3	3.56	153.3	3.52	164.4	3.55	154.5	3.58	151.7	3.50	150.9[32]
	N_2	3.67	143.5	3.74	129.2	3.66	144.7	3.72	130.3	3.76	127.4	3.68	127.6[32]
	O_2	3.56	152.5	3.65	129.4	3.54	155.6	3.63	132.0	3.67	130.7	3.57	133.8[32]
C_2H_2	He	3.34	42.0	3.63	25.7	3.32	42.4	3.62	26.0	3.69	27.6	3.34	46.5[2]
	Ar	3.79	94.5	4.01	75.2	3.80	94.6	4.01	75.3	4.06	75.3	3.72	170.1[2]
	N_2	4.23	41.7	4.35	38.2	4.23	43.1	4.36	39.5	4.40	37.0	3.90	143.9[2]
	O_2	3.67	150.2	3.94	99.5	3.63	154.1	3.90	101.8	4.01	95.6	3.79	150.9[2]
C_2H_4	He	3.49	42.1	3.68	30.6	3.47	42.5	3.67	30.7	3.70	32.3	3.39	47.3[32]
	Ar	3.81	149.9	3.95	124.7	3.79	150.9	3.94	125.1	3.98	124.3	3.77	172.9[32]
	N_2	4.06	101.4	4.18	88.7	4.05	102.9	4.17	90.0	4.24	85.4	3.94	146.3[32]
	O_2	3.78	173.4	3.96	131.1	3.74	179.3	3.93	133.5	4.02	128.4	3.83	153.4[32]
C_2H_6	He	3.63	41.5	3.76	33.8	3.62	41.8	3.75	33.9	3.80	33.8	3.40	52.9[32]
	Ar	3.86	196.8	3.97	169.2	3.84	200.1	3.95	170.3	4.00	169.1	3.78	193.6[32]
	N_2	4.03	163.2	4.15	138.9	4.01	165.4	4.13	140.0	4.21	134.1	3.95	163.7[32]
	O_2	3.88	192.4	4.03	154.0	3.84	199.2	3.99	157.1	4.08	153.7	3.84	171.7[32]
C_3H_8	He	3.87	48.5	4.10	35.2	3.86	48.9	4.09	35.3	4.16	34.3	3.71	55.3[32]
	Ar	4.11	222.3	4.31	172.0	4.07	227.7	4.28	173.7	4.36	169.2	4.08	202.3[32]
	N_2	4.29	180.4	4.49	141.0	4.25	184.0	4.46	142.5	4.58	130.8	4.26	171.2[32]
	O_2	4.13	220.2	4.36	162.2	4.07	229.8	4.31	166.2	4.46	154.8	4.15	179.5[32]
n-C_4H_{10}	He	4.05	54.8	4.57	29.5	4.03	55.5	4.55	29.5	4.71	27.6	3.79	63.8[32]
	Ar	4.31	245.8	4.79	145.0	4.23	254.3	4.72	144.9	4.90	137.9	4.17	233.4[32]
	N_2	4.50	196.2	4.95	120.9	4.43	201.5	4.90	121.3	5.12	103.5	4.34	197.4[32]
	O_2	4.33	245.5	4.82	140.0	4.22	259.9	4.74	141.5	5.00	124.1	4.23	207.0[32]
n-C_5H_{12}	He	4.22	59.6	5.09	23.6	4.18	60.6	5.07	23.4	5.30	21.6	3.93	69.6[32]
	Ar	4.48	264.0	5.31	117.2	4.35	276.5	5.22	113.8	5.48	108.5	4.31	254.4[32]
	N_2	4.68	208.8	5.46	99.1	4.58	216.1	5.40	97.3	5.72	80.4	4.48	215.2[32]
	O_2	4.51	265.2	5.33	114.5	4.34	285.3	5.22	112.3	5.58	97.8	4.37	225.7[32]
n-C_6H_{14}	He	4.35	64.2	5.64	18.7	4.30	65.4	5.62	18.3	5.96	15.7	4.25	65.7[2]
	Ar	4.62	281.1	5.85	93.6	4.44	296.0	5.74	87.6	6.13	80.7	4.62	240.3[2]
	N_2	4.83	220.5	6.01	80.0	4.69	229.2	5.92	76.7	6.35	61.6	4.80	203.3[2]
	O_2	4.64	283.7	5.88	92.2	4.42	307.8	5.74	87.0	6.21	75.7	4.69	213.1[2]
n-C_7H_{16}	He	4.45	68.0	6.22	15.0	4.39	69.3	6.19	14.6	6.54	12.7	4.15	79.7[32]
	Ar	4.74	295.7	6.43	75.8	4.51	312.3	6.30	67.8	6.72	64.7	4.52	291.2[32]
	N_2	4.95	230.4	6.58	65.2	4.79	240.0	6.49	60.5	6.99	47.7	4.70	246.3[32]
	O_2	4.76	299.1	6.45	74.9	4.49	325.9	6.29	67.3	6.85	59.4	4.59	258.3[32]
n-C_8H_{18}	He	4.54	71.3	6.80	12.3	4.47	72.7	6.76	11.8	7.19	10.0	4.99	58.3[2]
	Ar	4.83	308.6	7.01	62.2	4.57	325.5	6.86	53.7	7.35	51.4	5.37	213.2[2]
	N_2	5.05	239.3	7.15	53.8	4.86	249.2	7.05	48.6	7.61	38.6	5.54	180.4[2]
	O_2	4.85	313.0	7.02	61.8	4.55	341.0	6.85	53.4	7.46	48.2	5.43	189.1[2]
n-C_9H_{20}	He	4.61	74.2	7.34	10.5	4.53	75.6	7.30	10.0	7.78	8.3	5.44	52.1[2]
	Ar	4.91	320.1	7.55	53.1	4.63	337.1	7.40	44.3	7.95	42.7	5.82	190.6[2]
	N_2	5.14	247.3	7.70	45.9	4.93	257.4	7.59	40.5	8.23	31.7	5.99	161.2[2]
	O_2	4.93	325.5	7.57	52.9	4.60	354.6	7.38	43.9	8.07	39.2	5.88	169.0[2]
$C_{24}H_{12}$	He	5.01	143.8	6.53	37.4	4.59	165.3	6.47	31.3	6.73	36.3	5.37	100.1[25]
	Ar	5.65	243.0	6.90	116.2	5.45	249.7	6.86	106.8	7.08	110.1	5.75	365.9[25]
	N_2	6.34	81.2	7.22	65.6	6.46	87.5	7.23	72.8	7.37	58.9	5.92	309.5[25]
	O_2	5.51	443.3	6.80	155.2	4.42	618.7	6.55	97.1	7.04	137.1	5.81	324.6[25]

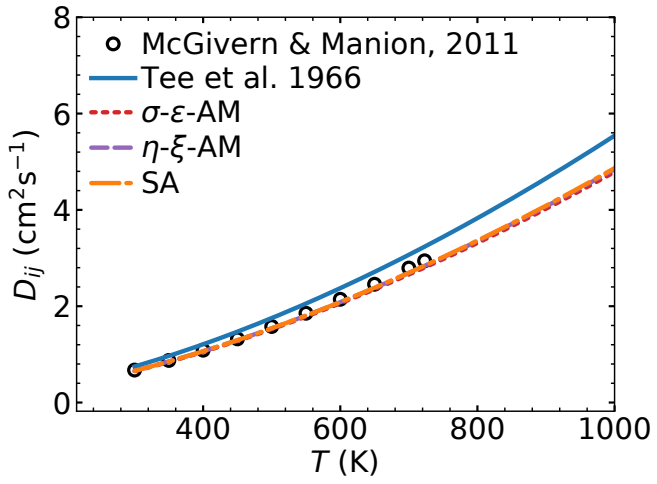


(a) CH_4, σ **(b)** CH_4, r_{min} **(c)** CH_4, ε

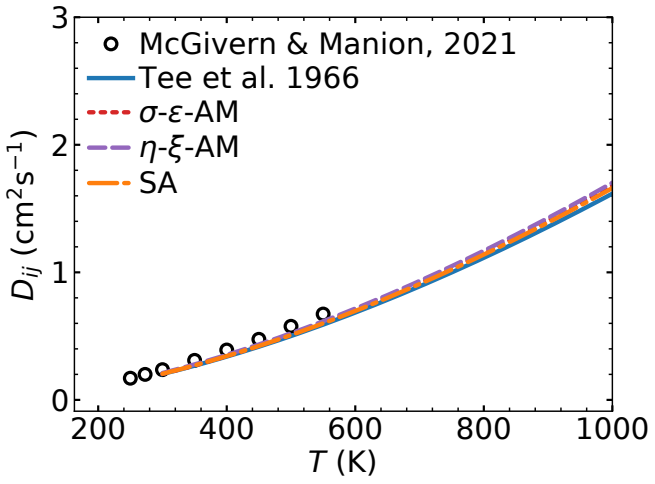
(d) n-C_9H_{20}, σ **(e)** n-C_9H_{20}, r_{min} **(f)** n-C_9H_{20}, ε

(g) $C_{24}H_{12}$, σ **(h)** $C_{24}H_{12}$, r_{min} **(i)** $C_{24}H_{12}$, ε

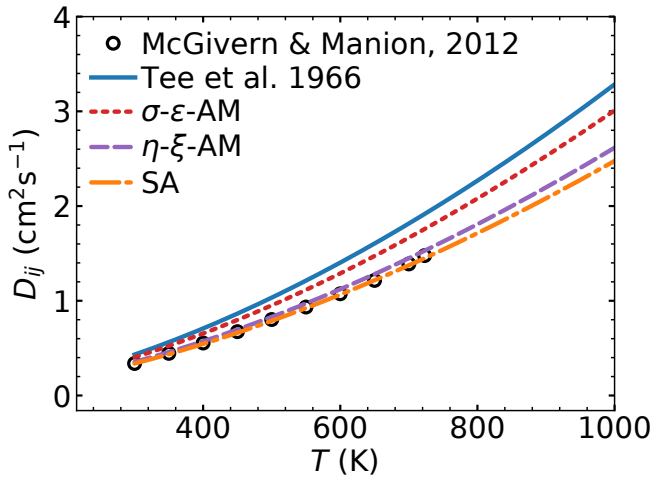
Fig. 3 Three-dimensional parametric surfaces of the collision diameter σ, well location r_{min}, and well depth ε by the developed iterative search method for a set of target-He pairs. The black dot at the origin, red dots, and blue dots represent the center of mass, carbon atoms, and hydrogen atoms, respectively. The distance from the origin to a point on the surface presents the magnitude of the quantity on that orientation. The molecular geometries in σ and r_{min} surfaces are of actual size, while those in ε surfaces are scaled up isotropically to better relate atoms with extrema on the parametric surfaces.



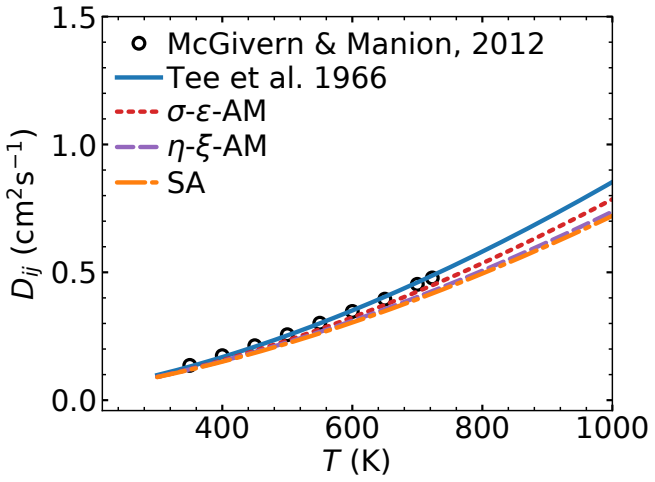
(a) CH_4 + He



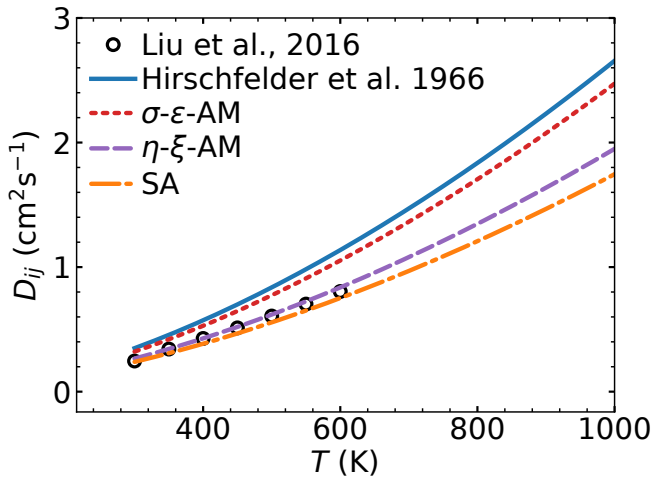
(b) CH_4 + N_2



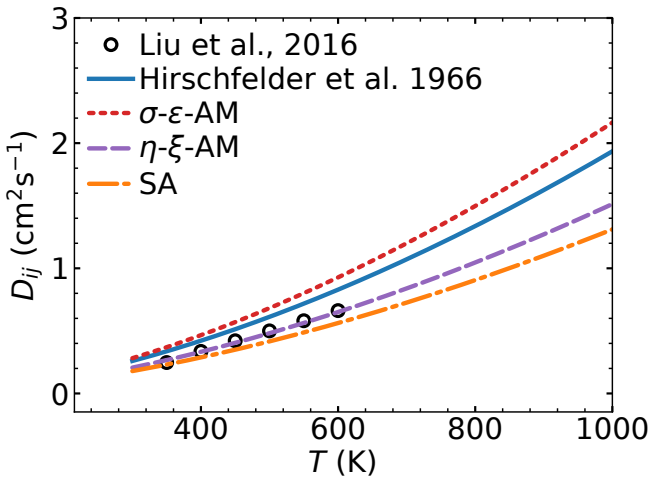
(c) n-C_4H_{10} + He



(d) n-C_4H_{10} + N_2



(e) n-C_6H_{14} + He



(f) n-C_8H_{18} + He

Fig. 4 Comparison of binary diffusion coefficients of several normal alkanes in He or N_2. Symbols are experimental diffusion data; solid lines are calculated by empirical parameters; dashed lines are calculated by Lennard-Jones parameters obtained via different averaging rules. The complete comparison for normal alkanes from C_1 to C_8 is provided in Supporting Information.

of the major macroscale outcomes resulting from molecular interactions.[2] Compared with the collected empirical Lennard-Jones parameters, the predicted parameters in this study are via first-principle calculations and do not involve the empirical Lorentz-Berthelot combining rules. Therefore, comparisons between binary diffusion coefficients measured by experiments and those calculated by the predicted Lennard-Jones parameters are able to directly reflect the effectiveness of the predicted parameters in the context of the Chapman-Enskog theory.

The Chapman-Enskog theory[20] is arguably the most successful theory for predicting transport coefficients of dilute gases based on the kinetic theory.[2,64,65] The binary diffusion coefficient can be formulated as

$$D_{ij} = \frac{3}{16}\frac{\sqrt{2\pi(k_B T)^3/m_{ij}}}{P\pi\sigma_{ij}^2\Omega^{(1,1)*}} \tag{18}$$

where T is temperature, P is pressure, m_{ij} is reduced mass of A + M, $\Omega^{(1,1)*}$ is a reduced collision integral that depends on the reduced temperature $T^* = k_B T/\varepsilon_{ij}$. After obtaining σ_{ij} and ε_{ij}, the calculation of D_{ij} in this study was facilitated by a computer code developed by Kee *et al.*[62].

Comparison of the binary diffusion coefficients calculated by the effective Lennard-Jones parameters of different averaging rules and by the collected empirical parameters with experimental diffusion data was conducted. Due to the availability of experimental diffusion data, the comparison is currently limited to normal alkanes. The employed experimental diffusion data are from McGivern and Manion[47,48,50] and Liu *et al.*[49], which cover diffusion coefficients at 1 atm pressure of normal alkanes in He or N_2 from C_1 to C_8 except the *n*-C_7H_{16}. The present dataset represents by far the latest and probably the most complete experimental diffusion data on normal alkanes. The complete comparison is shown in 12 plots and is provided in the Supporting Information. Because diffusion coefficients predicted by effective parameters related to σ-ε-BW and η-ξ-BW rules are almost identical to those related to σ-ε-AM and η-ξ-AM rules, respectively, diffusion coefficient curves of the former are omitted in the plots. The plots for CH_4, *n*-C_4H_{10}, *n*-C_6H_{14}, and *n*-C_8H_{18} are shown in Fig. 4 to better facilitate discussion. Note that only He related experimental data are available for *n*-C_6H_{14} and *n*-C_8H_{18}.

As can be seen, diffusion coefficients calculated by the empirical Lennard-Jones parameters from Tee *et al.*[32] and Hirschfelder *et al.*[2] agree well with the experimental diffusion data for bath gas N_2 but have large deviations for bath gas He in most cases. A similar behaviour can also be observed for the conventional σ-ε-AM rule. For example, the deviations for results of empirical parameters and σ-ε-AM rule at 600 K are about 41% and 31%, respectively, in the *n*-C_6H_{14} + He case (Fig. 4e) and are about 25% and 40%, respectively, in the *n*-C_8H_{18} + He case (Fig. 4f). These marked deviations agree with the results of Liu *et al.*[49], which motivated their development of an alternative theoretical model for binary diffusion coefficients of long-chain molecules using gas-kinetic theory analysis for slender bodies.[49]

Compared with the σ-ε-AM rule, the traditional spherical average (labeled as "SA" in Fig. 4) performs fairly well for the examined 12 cases covering 7 normal alkanes and bath gases He and N_2. This finding at some level contradicts that made by Jasper and Miller[43]. The main reason for this contradiction may be attributed to the different choices of reference data. Jasper and Miller[43] mainly used the empirical parameters from Tee *et al.*[32] and Hirschfelder *et al.*[2] as the reference data and also used the empirical Lorentz-Berthelot combining rules for evaluation. As discussed earlier, the current use of the contemporary experimental diffusion data from McGivern, Manion and co-workers[47–50] as reference data has several advantages over the empirical parameters, such as avoiding the possible non-uniqueness and uncertainties in deriving the empirical Lennard-Jones parameters via Chapman-Enskog relations as well as the Lorentz-Berthelot combining rules for inferring binary Lennard-Jones parameters. This result shows the importance of consistent and contemporary experimental transport data for verifying theoretical and numerical models at the meso- and micro-scales. The overall better performance of the spherical average over the σ-ε-AM rule indicates that averaging energies may better capture the intermolecular interaction statistics than directly averaging the collision diameters and well depths does.

When comparing the spherical average (SA) with the newly proposed η-ξ-AM rule, the maximum errors both happen at the highest measured temperatures. Among all the examined 12 cases, the maximum in 12 maximum errors by the spherical average (SA) and the η-ξ-AM rule are 19% and 12%, respectively. The average of 12 maximum errors by the spherical average (SA) and the η-ξ-AM rule are 8% and 6%, respectively. In addition, the accuracy of the spherical average deteriorates for very long-chain molecules such that its maximum errors for *n*-C_6H_{14} and *n*-C_8H_{18} become about 7% and 15%, respectively. Meanwhile, the maximum errors of the η-ξ-AM rule for *n*-C_6H_{14} and *n*-C_8H_{18} are only about 4% and 2%, respectively. As suggested by Jasper and Miller[43], the rapid increase of errors from *n*-C_6H_{14} to *n*-C_8H_{18} by the spherical average can be due to the increased portion of mixing repulsive walls with attractive regions during spherically averaging potentials around highly anisotropic molecules, which affects the subsequently derived effective parameters. The proposed η-ξ-AM rule in this study avoids this issue by explicitly averaging the anisotropic energy scales.

These quantitative results suggest that the η-ξ-AM rule can produce the most accurate results among the evaluated averaging rules, and this rule can maintain good consistency in accuracy for very long-chain molecules, which indicates that the η-ξ-AM rule built on averaging the characteristic repulsive and dispersive energies can be more effective in representing the statistical outcomes of frequent molecular rotations and collisions.

4 Conclusions

In this work, the first-principle determination of effective Lennard-Jones parameters for polyatomic molecules interacting with bath gas molecules has been investigated. To efficiently and robustly find orientation-dependent collision diameters and well depths on intermolecular potential energy surfaces, an iterative search algorithm built on projection preprocessing and modified bisection strategy is developed and validated. To quantitatively

depict molecular anisotropy, three-dimensional parametric surfaces of Lennard-Jones parameters are constructed. To derive the effective Lennard-Jones parameters, a new orientation-averaging rule based on characteristic variables, the η-ξ-AM rule, is proposed and evaluated.

Effective Lennard-Jones parameters for twelve saturated and unsaturated hydrocarbons with varying molecular shapes, including long-chain and planar ones, interacting with four bath gases He, Ar, N_2, and O_2 are reported. By comprehensive quantitative comparisons against empirical Lennard-Jones parameters, various averaging methods, and latest experimental diffusion data, it is found that the proposed η-ξ-AM rule can produce the most accurate results among the evaluated averaging rules and resolves the observed inconsistencies in using the Chapman-Enskog theory for polyatomic molecules, which is a problem of theoretical interest and practical importance.

Compared with previous averaging methods, the η-ξ-AM rule explicitly averages the characteristic repulsive and dispersive energies, which could better capture intermolecular interactions than directly averaging collision diameters and well depths does. Meanwhile, this rule avoids mixing the repulsive walls with attractive regions in the averaging process for highly anisotropic molecules, which is a potential issue in the spherically averaging potential energies approach. By introducing characteristic variables η and ξ, the current work shows a new route to determining effective Lennard-Jones parameters for polyatomic molecules.

The historical development of averaging rules suggests that further advancement of the current and related topics may rely on the further acquisition of consistent and contemporary experimental transport data that can be used to verify theoretical and numerical models at the meso- and micro-scales.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Financial support of this work was provided by the National Natural Science Foundation of China (Grant No. 51976099). Supercomputing time on ARCHER is provided by the "UK Consortium on Mesoscale Engineering Sciences (UKCOMES)" under the UK Engineering and Physical Sciences Research Council Grant No. EP/R029598/1.

References

1 E. Mason and L. Monchick, *J. Chem. Phys.*, 1962, **36**, 2746–2757.

2 J. O. Hirschfelder, C. F. Curtiss, R. B. Bird and M. G. Mayer, *Molecular Theory of Gases and Liquids*, Wiley New York, 1964, vol. 165.

3 R. J. Kee, G. Dixon-Lewis, J. Warnatz, M. E. Coltrin and J. A. Miller, *Sandia National Laboratories Report SAND86-8246*, 1986, **13**, 80401–1887.

4 S. Stephan, J. Staubach and H. Hasse, *Fluid Phase Equilib.*, 2020, **523**, 112772.

5 G. Rutkai, M. Thol, R. Span and J. Vrabec, *Mol. Phys.*, 2017, **115**, 1104–1121.

6 J. A. Miller, R. Sivaramakrishnan, Y. Tao, C. F. Goldsmith, M. P. Burke, A. W. Jasper, N. Hansen, N. J. Labbe, P. Glarborg and J. Zádor, *Prog. Energy Combust. Sci.*, 2021, **83**, 100886.

7 J. E. Jones, *Proc. R. Soc. London, Ser. A*, 1924, **106**, 463–477.

8 S. Stephan, M. T. Horsch, J. Vrabec and H. Hasse, *Mol. Simul.*, 2019, **45**, 806–814.

9 B. L. Eggimann, A. J. Sunnarborg, H. D. Stern, A. P. Bliss and J. I. Siepmann, *Mol. Simul.*, 2014, **40**, 101–105.

10 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.

11 R. L. Akkermans, N. A. Spenley and S. H. Robertson, *Mol. Simul.*, 2021, **47**, 540–551.

12 N. J. Brown, L. A. Bastien and P. N. Price, *Prog. Energy Combust. Sci.*, 2011, **37**, 565–582.

13 S. Stephan, M. Thol, J. Vrabec and H. Hasse, *J. Chem. Inf. Model.*, 2019, **59**, 4248–4265.

14 H. Wang, X. You, M. A. Blitz, M. J. Pilling and S. H. Robertson, *Phys. Chem. Chem. Phys.*, 2017, **19**, 11064–11074.

15 A. Laricchiuta, G. Colonna, D. Bruno, R. Celiberto, C. Gorse, F. Pirani and M. Capitelli, *Chem. Phys. Lett.*, 2007, **445**, 133–139.

16 W. Forst, *Theory of Unimolecular Reactions*, Academic Press, New York, 1973.

17 S. C. Gilbert, R. G.and Smith, *Theory of Unimolecular and Recombination Reactions*, Blackwell Science, Oxford, 1990.

18 Y. Georgievskii, J. A. Miller, M. P. Burke and S. J. Klippenstein, *J. Phys. Chem. A*, 2013, **117**, 12146–12154.

19 D. R. Glowacki, C.-H. Liang, C. Morley, M. J. Pilling and S. H. Robertson, *J. Phys. Chem. A*, 2012, **116**, 9545–9560.

20 S. Chapman, T. G. Cowling and D. Burnett, *The Mathematical Theory of Non-uniform Gases*, Cambridge University Press, 1990.

21 M. J. Pilling and S. H. Robertson, *Annu. Rev. Phys. Chem.*, 2003, **54**, 245–275.

22 A. M. Mebel, Y. Georgievskii, A. W. Jasper and S. J. Klippenstein, *Faraday Discuss.*, 2016, **195**, 637–670.

23 E. N. Fuller, P. D. Schettler and J. C. Giddings, *Ind. Eng. Chem.*, 1966, **58**, 18–27.

24 W. Wakeham and D. Slater, *J. Phys. B*, 1973, **6**, 886.

25 H. Wang and M. Frenklach, *Combust. Flame*, 1994, **96**, 163–170.

26 F. Pirani, S. Brizi, L. F. Roncaratti, P. Casavecchia, D. Cappelletti and F. Vecchiocattivi, *Phys. Chem. Chem. Phys.*, 2008, **10**, 5489–5503.

27 P. Middha, B. Yang and H. Wang, *Proc. Combust. Inst.*, 2002, **29**, 1361–1369.

28 M. Arias-Zugasti, P. L. Garcia-Ybarra and J. L. Castillo, *Combust. Flame*, 2016, **163**, 540–556.

29 M. Andac and F. Egolfopoulos, *Proc. Combust. Inst.*, 2007, **31**, 1165–1172.

30 A. Holley, X. You, E. Dames, H. Wang and F. N. Egolfopoulos, *Proc. Combust. Inst.*, 2009, **32**, 1157–1163.

31 C. Ji, E. Dames, Y. L. Wang, H. Wang and F. N. Egolfopoulos, *Combust. Flame*, 2010, **157**, 277–287.

32 L. S. Tee, S. Gotoh and W. E. Stewart, *Ind. Eng. Chem. Fundam.*, 1966, **5**, 356–363.

33 P. Paul and J. Warnatz, *Proc. Combust. Inst.*, 1998, **27**, 495–504.

34 J. Yarnell, M. Katz, R. G. Wenzel and S. Koenig, *Phys. Rev. A*, 1973, **7**, 2130.

35 A. Burgmans, J. Farrar and Y. Lee, *J. Chem. Phys.*, 1976, **64**, 1345–1350.

36 Y. T. Lee, *Science*, 1987, **236**, 793–798.

37 G. E. Ewing, *Canadian J. Phys.*, 1976, **54**, 487–504.

38 J. M. Hutson, *Annu. Rev. Phys. Chem.*, 1990, **41**, 123–154.

39 E. Zunzunegui-Bru, E. Gruber, S. Bergmeister, M. Meyer, F. Zappa, M. Bartolomei, F. Pirani, P. Villarreal, T. González-Lezana and P. Scheier, *Phys. Chem. Chem. Phys.*, 2021.

40 S. Bechtel, B. Bayer, T. Vidaković-Koch, A. Wiser, H. Vogel and K. Sundmacher, *Int. J. Heat Mass Transf.*, 2020.

41 V. Bernshtein and I. Oref, *J. Phys. Chem. A*, 2000, **104**, 706–711.

42 M. Shadman, S. Yeganegi and F. Ziaie, *Chem. Phys. Lett.*, 2009, **467**, 237–242.

43 A. W. Jasper and J. A. Miller, *Combust. Flame*, 2014, **161**, 101–110.

44 L. Monchick and S. Green, *J. Chem. Phys.*, 1975, **63**, 2000–2009.

45 K. F. Lim, *J. Chem. Phys.*, 1994, **100**, 7385–7399.

46 G. Lendvay, *Unimolecular Kinetics*, Elsevier, 2019, vol. 43, pp. 109–272.

47 W. S. McGivern and J. A. Manion, *J. Chromatogr. A*, 2011, **1218**, 8432–8442.

48 W. S. McGivern and J. A. Manion, *Combust. Flame*, 2012, **159**, 3021–3026.

49 C. Liu, W. S. McGivern, J. A. Manion and H. Wang, *J. Phys. Chem. A*, 2016, **120**, 8065–8074.

50 W. S. McGivern and J. A. Manion, *J. Chem. Eng. Data*, 2021.

51 H. Lorentz, *Annalen der physik*, 1881, **248**, 127–136.

52 C. Liu, R. Zhao, R. Xu, F. N. Egolfopoulos and H. Wang, *Proc. Combust. Inst.*, 2017, **36**, 1523–1530.

53 E. A. Mason, *J. Chem. Phys.*, 1954, **22**, 169–186.

54 K. Tang and J. P. Toennies, *J. Chem. Phys.*, 1984, **80**, 3726–3741.

55 X. Wang, S. Ramírez-Hinestrosa, J. Dobnikar and D. Frenkel, *Phys. Chem. Chem. Phys.*, 2020, **22**, 10624–10633.

56 J. A. Barker and D. Henderson, *Rev. Mod. Phys.*, 1976, **48**, 587.

57 A. W. Jasper and J. A. Miller, *A code for calculating Lennard–Jones parameters from detailed intermolecular potentials via one-dimensional minimizaiton*, 2014, `https://tcg.cse.anl.gov/papr/codes/onedmin.html`.

58 R. D. Johnson III *et al.*, *NIST 101. Computational chemistry comparison and benchmark database*, 2020, `http://cccbdb.nist.gov/`.

59 S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

60 A. W. Jasper and J. A. Miller, *J. Phys. Chem. A*, 2011, **115**, 6438–6455.

61 B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, CRC press, 1994.

62 R. J. Kee, F. M. Rupley and J. A. Miller, *Sandia National Laboratories Report SAND-89-8009*, 1989.

63 L. A. Bastien, P. N. Price and N. J. Brown, *Int. J. Chem. Kinet.*, 2010, **42**, 713–723.

64 A. A. Medvedev, V. V. Meshkov, A. V. Stolyarov and M. C. Heaven, *Phys. Chem. Chem. Phys.*, 2018, **20**, 25974–25982.

65 F. Sharipov and V. J. Benites, *Phys. Chem. Chem. Phys.*, 2021, **23**, 16664–16674.