An Investigation of Molecular Dynamics for Simple Liquids

Jonathan N.F.W Utterson
Merton College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary 2023

"All that is gold does not glitter, not all those who wander are lost" -

J.R.R. Tolkien

Acknowledgements

My time at Oxford has been a journey of constant missteps, mishaps and misgivings - yet I have no doubt that when I look back at this time, I will miss it.

I thank my supervisor, Radek Erban, for supervising me these past years - allowing me the freedom to explore, and steering when I wander into particularly treacherous territories. He has pushed me further than I thought possible, and patiently motivated me to try for more.

The WCMB has been a more than welcoming group, headed by an immensely supportive Phillip Maini - thank you for all of the conversations you have taken the time to have with me.

A big thanks goes to Karen, Mike and Tom: who keep me tethered, especially in moments when life feels like it is getting away from me. To Jenny, you have made all of this possible and you continue to inspire me, I will forever strive to be as supportive as you have been in all aspects of my life.

To my friends and family, you mean more to me than you could possibly know.

Abstract

This thesis contains work expanding the theoretical understanding of molecular dynamics used to aid the study of simple liquids. It does so by focusing on investigating forces, which govern the dynamics of many-body systems. We loosely address three questions: *How can we categorise force distributions? What can we gauge from force data? When do forces obey Newton's third law?*

The first of these questions is addressed using statistical mechanics to derive standardised moments of the force distribution for a simple Lennard-Jones liquid in both 1d and 3d with the aid of molecular dynamics.

To answer the second question, we introduce the notions of force spaces and configurations spaces, and look at equivalence of these. We begin the investigation using the harmonic potential, and develop homotopy continuation methods for non-linear forces like Lennard-Jones. Convergent behaviour and limitations are explored for many-body systems, and a general two-body direct inversion is developed and implemented.

The final question is entrenched in classical potential theory, and approached through work focusing on understanding the functional dependence of the interatomic potential. We develop theorems and provide corresponding constructive proofs concluding that potentials which obey certain symmetries can be described by distances, as opposed to positions. This enables us to understand when forces display reciprocity.

Contents

List of Figures

List of Tables

List of Abbreviations

MD	Molecular dynamics
NHCM	Newton homotopy continuation method
LJ	Lennard-Jones
CG	Coarse grained
SDE	Stochastic differential equation
NHL	Nosé-Hoover-Langevin
AA	All-atom
PMF	Potential of mean force
MS-CG	Multi-scale coarse graining
PBC	Periodic boundary condition
CPT	Central potential theorem

List of Symbols

N	Number of particles or atoms in a system
V	Volume of accessible position space
T	Temperature of system
E	Energy of system
$\mathbf{M}$	Diagonal matrix of atom masses
m_i	The mass of a single atom indexed i
$\mathbf{q}$	Vector of all positions of atoms
$\mathbf{q}_i$	Position vector for i^{th} atom
$\mathbf{p}$	Vector of all momenta of atoms
$\mathbf{p}_i$	Momentum vector for i^{th} atom
Δt	Time-step of a numerical integrator
ξ	Nosé-Hoover auxiliary variable
μ	Nosé-Hoover relaxation parameter
N_d	Number of spatial degrees of freedom
k_B	Boltzmann constant
β	$1/k_BT$
$\mathcal{H}$	Hamiltonian
r	Distance between two interacting bodies
n	Number density of system of atoms
F_{helm}	Helmholtz free energy
$\tilde{\mathbf{q}}$	Vector of all positions of CG sites
$\tilde{\mathbf{q}}_i$	Position vector for the i^{th} CG site
N_{CG}	Number of CG sites in a system
U_{CG}	Potential mean force for CG system
$\mathcal{M}_{\tilde{\mathbf{q}}_I}(\mathbf{q})$	Linear mapping from atomic positions to I^{th} CG site

$\mathcal{M}_{\tilde{\mathbf{q}}}^{N_{CG}}(\mathbf{q})$	Linear mapping from atomic positions to all CG sites
δ	Dirac delta function
$\mathbf{F}_i$	Total force on the i^{th} atom
$\mathcal{X}$	Configuration space
$\mathcal{F}$	Force space
F	Mapping from configuration space to force space
$\mathbf{F}$	Vector of total forces on all atoms
$\underline{\underline{A}}$	Matrix describing linear mapping F
$\mathbf{Q}$	A vector field used to set up initial equations in NHCM
$\mathbf{P}$	A vector field used to set up target equations in NHCM
$\mathbf{H}$	Linear homotopy function
γ	Step size used in homotopy continuation methods
$\tilde{n}$	Number of atoms in a subsystem of N atoms
$\mathbf{F}_{ij}$	Pairwise force on atom i from atom j
r_{ij}	Pairwise interatomic distance between atoms i and j
α_k	The k^{th} standardised moment of the force distribution
U_{ij}	Pairwise potential between atoms i and j
$\mathcal{Z}_N$	Canonical partition function for N atoms
h	Plank's constant
r^*	Potential equilibrium distance
$\Omega_{\mathbf{q}}$	Position space for many-body systems
$\Omega_{\mathbf{p}}$	Momentum space for many-body systems
L	Box width for simulation domain
X	A general statistical quantity
F_i	The i^{th} component of force on a tagged atom
$F_{j,i}$	The i^{th} component of force on the j^{th} atom
f_k	The k^{th} marginalised moment of the force distribution

q_0	Fixed initial position in 1d
x	Position in 1d
δL	Infinitesimal change in box width
c	Cut-off parameter in 1d
$\tilde{\gamma}$	Friction parameter in Langevin dynamics
γ_c	Cut-off parameter in 3d
$< \cdot >$	Thermodynamic average
$\lfloor \cdot \rceil$	Rounds to nearest integer
L_0	Initial box width
t_{sim}	Total simulation time
T_c	Critical temperature for clustering
$R(t)$	White noise
k	(Harmonic spring) constant
k_{ij}	Pairwise harmonic spring constant between atoms i and j
$\mathbf{\Delta}_{ji}$	Displacement vector between atoms i and j
$\|\cdot\|'_\infty$	Scaled L^∞-norm of a vector
$\|\cdot\|_\infty$	L^∞-norm of a vector
$\mathbf{q}^0$	Initial cubic lattice positions
$\bar{\mathbf{q}}$	Positions after translation to middle of domain
$\mathbf{\Delta}_\gamma$	Vector of all single jump displacements
n_1	Number of Newton's methods implemented in NHCM
n_2	Number of iterations per Newton's method
$\Delta\gamma$	Step size for linear homotopy function
J	Jacobian of the linear homotopy function
$\mathbf{\Delta}^{i,j}$	The vector of all single jump displacements for the i^{th} Newton's method used for calculations on the j^{th} iteration
$\mathbb{I}_{3x3}$	The 3x3 identity matrix

$\underline{\underline{\mathbf{F}}}$	Matrix of total forces
n_t	Number of time steps recording force data
$\Theta(x)$	The Heaviside function
$\mathbf{rd}_i$	The vector of distances corresponding to the i^{th} solution of a polynomial
$\mathbf{r}$	A vector of distances
U_n	General n-body potential
Ψ_n	General n-body distance dependent potential
$\phi(\mathbf{r})$	A general distance dependent potential
R	A rotation in $SO(3)$
Q	A reflection

Chapter 1

Introduction

Forces dictate the dynamics of many-body systems from the quantum to the cosmic scale. In this thesis, we focus on exploring how to categorise force distributions on the atomic scale. This is motivated by the growing need to develop lower resolution models which accurately represent the original system. We then consider more generally what the knowledge of forces tells us, and the consequences that physical symmetries such as translation and rotation have on describing them. These ideas are fundamental to understanding how we should view an interacting many-body system, which has amassed interest in all areas of science over the last few millennia.

Chapter 1 introduces the thesis by laying the foundation for the work. In Section 1.1, we briefly outline the key ideas behind molecular dynamics (MD), which are pertinent to work in Chapters 2, 3 and 4. We discuss how MD came to be one of the most ubiquitous techniques in science, describe where development of this field is heading, and the problems it faces to this day.

Section 1.2 focuses on force distributions, highlighting milestones in approaches to studying these distributions for many-body systems. This sets the tone for Chapter 2 where we investigate how a force distribution can depend on intrinsic MD parameters.

In Section 1.3 we expound on the concept of coarse graining, and point out several

methods commonly used in the MD community - though focus is given to the method of force-matching which motivates the novel idea of force inversion introduced in Section 1.4, and developed in Chapter 3. A key tool used in the implementation of force inversion is the Newton homotopy continuation method (NHCM), for which we provide a background and review in Section 1.5.

In the penultimate Section 1.6, we discuss classical interatomic potential theory focusing on the applications to non-reciprocal interactions. This promotes the new theorems introduced in Chapter 4, which provide a more rigorous footing to this field of study.

Finally, we conclude the introduction in Section 1.7 by giving the overall structure of the thesis.

1.1 Molecular dynamics

Molecular dynamics (MD) is the versatile computational method through which chemical, biological and physical systems can be described. It is a framework in which we first quantise a system in terms of unit substructures. In MD these will typically be particles, atoms and molecules; though more complicated structures such as groups of chained atoms can be considered in protein folding applications [16]. In MD, we look at atomistic scales typically characterised by ångström ($1\text{Å} = 10^{-10}\text{m}$) length scales. Crucially, we assume dynamics are modelled under the classical regime with Newton's equations of motion (1.1), whereas subatomic scales are best described by quantum mechanics, and larger scales are governed by continuum mechanics; which is neatly summarised by Berendsen's hierarchy [21]. An illustration of modelling hierarchy is given in Figure 1.1.

Typically, an interatomic potential through which these structures interact is prescribed, and this dictates the forces and therefore dynamics in the system of atoms. Usually these will govern Van der Waals forces between atoms for the scales we are

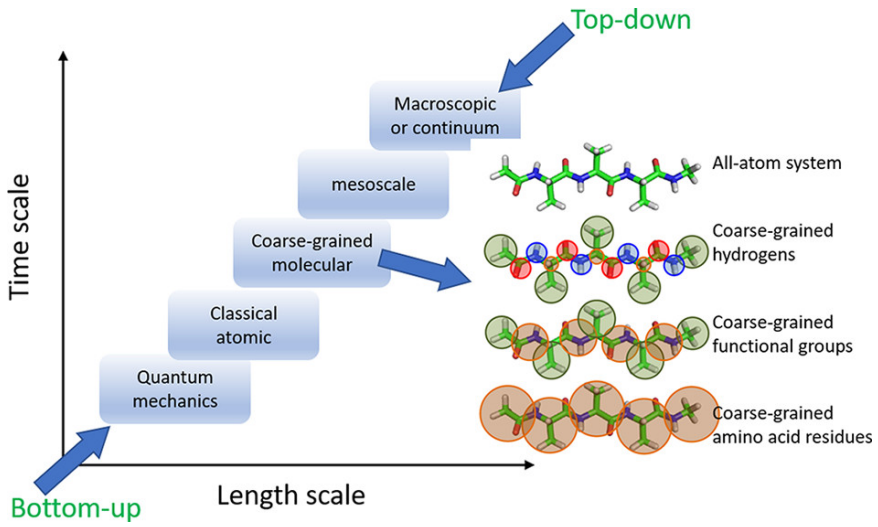


Figure 1.1: *This schematic is reprinted with permission from [234] which shows a basic hierarchy of modelling. Bottom-up modelling relies on developing low resolution models from higher resolutions, and top-down modelling uses phenomenological observations from the low resolution models to inform development of higher resolution models. The process of coarse graining an all atom model of a polypeptide, polyalanine, is shown on the right.*

interested in. One must judiciously choose this potential to best capture the physics of the system. For example: interactions can be approximated by the harmonic potential as interatomic springs connecting structures. This is used in the Rouse model [184] and is implemented ubiquitously in polymer physics [96]. We present an array of common pairwise potentials in Table 1.1, whilst non-pairwise potentials are discussed in Section 1.6.

Throughout the work in Chapter 2 and Chapter 3, we will be utilising the Lennard-Jones (LJ) potential [128]. This potential is a popular choice because it captures the dynamics of simple fluids well [205], it has a longstanding history of literature and large data repositories associated with it [206], and in particular (for Chapter 3) it provokes interesting questions surrounding the two attractive regimes the potential admits. The LJ potential is a simple two parameter continuous function, which does not suffer the same deficiencies as square-well/shoulder potentials used in basic modelling of simple liquids [110]. The discontinuities of the shoulder potentials lead to infinite forces that must be carefully time-stepped around whereas the LJ potential

Model	Description
Lennard-Jones (LJ) [128]	Short-ranged due to the fast decaying nature of the power law (usually 12-6 potential). Short-range Pauli forces and long-range attractive forces are captured.
Morse [162]	Used for modelling quantum systems and calculating bond dissociation.
Buckingham [30]	Modification of the LJ potential by utilising an exponential form for Pauli repulsion but the same form for London dispersion.
Yukawa [236]	An altered form of the Coulombic potential which captures electron screening.
Gay-Berne [97]	An anisotropic form of the LJ potential arising from the spherical to ellipsoidal body approximation.

Table 1.1: *A variety of commonly used interatomic potentials.*

is smooth. This is not to say that the LJ potential itself is without fault, indeed it has come under rightful scrutiny due to the popularity it has amassed. For example, the London dispersion effects are sometimes inappropriate [165, 66], which plays into the more general feature that this is a potential best suited for inert systems [84], where its roots are firmly based with the seminal study of liquid argon in the 1960s [177]. Other common potentials in MD are given in Table 1.1 for completeness, each has a unique contribution in the development of the field.

In general, MD is an all encompassing method which has grown from modelling simple fluids, to probing quantum structures to drug discovery, the review and application of which has received much attention with details best communicated in recent papers [3, 119, 141]. MD fundamentally relies on analysing equations of motion to create trajectories: these are the time evolutions for atoms (or particles, molecules *etc*) in the position-momentum space.

In this work we frame MD in a classical setting. This is termed molecular mechanics, whereby prescribing a semi-empirical potential under the Born-Oppenheimer approximation (which detaches the nucleus from electron motion) allows us to largely describe the classical motion of atoms using equations of motion admitted by New-

ton's second law [154]. We will mainly be concerned in Chapters 2, 3 and 4 with these equations introduced below. For a system of N atoms with positions $\mathbf{q} = (\mathbf{q}_1, \ldots, \mathbf{q}_N) \in \mathbb{R}^{3N}$, momenta $\mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_N) \in \mathbb{R}^{3N}$, masses given by a mass matrix $\mathbf{M} = \text{diag}(m_1, m_2, ..., m_N)$, interacting via a potential $U(\mathbf{q}) : \mathbb{R}^{3N} \to \mathbb{R}$:

$$\mathbf{M}\ddot{\mathbf{q}} = \mathbf{F}(\mathbf{q}) = -\nabla U(\mathbf{q}), \tag{1.1}$$

where the vector of forces is denoted $\mathbf{F} = (\mathbf{F}_1, \ldots, \mathbf{F}_N) \in \mathbb{R}^{3N}$.

It is important to note that one can describe the system of equations in (1.1) as a coupled set of first order ODEs in the following way:

$$\dot{\mathbf{q}} = \mathbf{M}^{-1}\mathbf{p} \tag{1.2a}$$

$$\dot{\mathbf{p}} = \mathbf{F}(\mathbf{q}) = -\nabla U(\mathbf{q}). \tag{1.2b}$$

Due to the connection to Hamilton's formulation, the set of equations (1.2) are extremely versatile, and many numerical schemes have been implemented to integrate them [143].

Other deterministic equations of motion include the PDE description given by Schrödinger's equation for quantum scales. This is outside of the remit of molecular mechanics, but a hybridised approach forms quantum mechanics-molecular mechanics modelling [195]. Systems are treated ab-initio where electron structures are calculated 'on-the-fly' [139]. Though we do not work in this direction, it was one of the most pivotal achievements to use MD to probe quantum scales with the pioneers Warshel, Levitt and Karplus being awarded the Nobel prize in chemistry 2013 [168] for their vital contribution to multiscale-modelling, especially for their applications to modelling enzymatic reactions [227].

MD is a simulation technique which has been invaluable to the study of biophysics. Karplus provides an excellent overview of MD [131] and discusses the three main areas of interest in biomolecular simulation: sampling configuration spaces (the domain of all possible $\mathbf{q}$), simulating near equilibrium, or obtaining dynamical information. MD in particular is best at *dynamically* analysing these many-body systems. Note however that non-dynamic approaches such as Monte Carlo methods are particularly important in configuration sampling [170], especially in cases where a potential admits barriers that are hard to overcome with usual MD. This is a double-edged sword because there are some systems where high potentials lead to an overwhelming rejection rate meaning the sampling is biased [6]. Hybrid Monte Carlo methods often assuage these issues by making use of dynamical knowledge with a guidance Hamiltonian [76].

Stochastic differential equations (SDEs) can also model many-body systems effectively, leading to stochastic equations of motion which capture randomness. A large class of equations are called the Ornstein-Uhlenbeck equations [99, 144]:

$$d\mathbf{p} = -\tilde{\gamma}\mathbf{p}\,dt + \sqrt{2\tilde{\gamma}k_BT}\,\mathbf{M}^{1/2}\,d\mathbf{W}, \tag{1.3}$$

which are described by a friction parameter $\tilde{\gamma}$, an infinitesimal time step dt, Boltzmann's constant k_B and temperature T. The vector $\mathbf{W} = (W_1, \ldots, W_{3N})$, where W_i is a Wiener process, means that dW_i captures the stochasticity at play as a Gaussian noise.

Equations (1.3) model non-interacting atoms; for interacting systems, a potential that gives rise to interatomic forces also contributes to the equations of motion, which defines Langevin dynamics [231]. The full set of equations are as follows:

$$d\mathbf{q} = \mathbf{M}^{-1}\mathbf{p}\,dt \tag{1.4a}$$

$$d\mathbf{p} = -\nabla U(\mathbf{q})\,dt - \tilde{\gamma}\mathbf{p}\,dt + \sqrt{2\tilde{\gamma}k_BT}\,\mathbf{M}^{1/2}\,d\mathbf{W}. \tag{1.4b}$$

The ergodic hypothesis [200] is vital in the utility of MD: it states that calculating temporal averages of MD trajectories, or averaging over enough sampled configurations *should* give the same expected values as thermodynamic averages [160]. In general this is difficult to prove with deterministic systems: in fact it has been shown to not hold for certain simple systems [216]. A common fix is to implement stochastic noise to perturb trajectories into the whole configuration space with equations (1.4). In Chapter 2, we utilise Langevin dynamics to simulate a 1d LJ system to ensure such sampling is robust.

Time correlation functions [111] are defined as the thermodynamic average of the product of two dynamical variables, used initially in simple liquid modelling [238], but more recently for developing coarse-grained (CG) models discussed in Section 1.3. Green-Kubo relations help determine transport coefficients with appropriate knowledge of these time correlation functions [194]: a simple example is outlined in [26] where the velocity autocorrelation function is related to the self diffusion coefficient.

The force autocorrelation function in particular can be used to calculate the self diffusion coefficient of a molecule in a simple liquid [113] and helps us understand friction in general [173], especially in the context of random motion [52].

To sample from a specific ensemble (such as the canonical or isobaric ensembles), unaltered deterministic equations of motion are not well equipped to maintain the temperature T of the system. As such, temperature thermostats are used in place of a thermal bath, which allow energy exchange in order to simulate at an average target temperature. A common example is the Nosé-Hoover thermostat [167]. However, it is worth noting that the implementation of the Nosé-Hoover thermostat to non-ergodic systems such as a 1d harmonic oscillator sometimes remains non-ergodic [216]. One can potentially improve the ergodicity with the introduction of a thermal bath made of many Nosé-Hoover chains [156]. For more guaranteed cor-

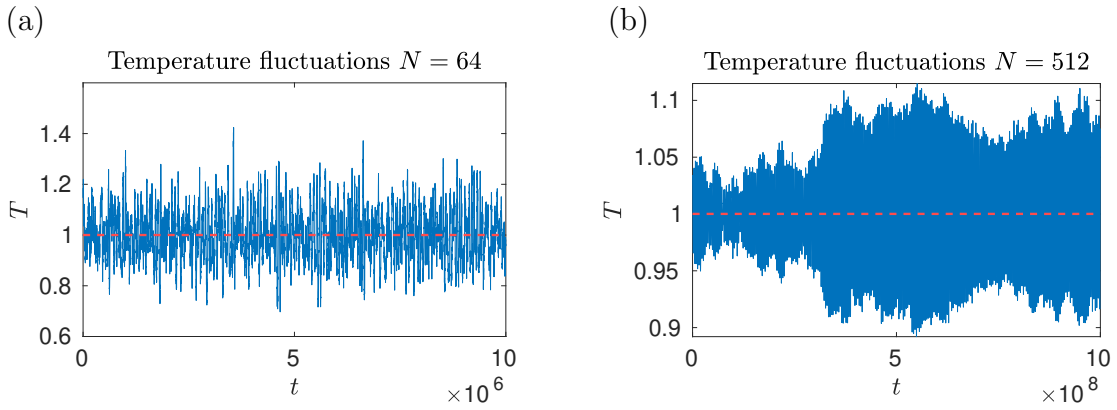


Figure 1.2: *Here we illustrate the typical temperature fluctuations for a LJ fluid seen in Chapter* 2*, with number density* $n = 1$*, parameters* $\tilde{\gamma} = \mu = 1$*, while maintaining a target temperature of* $T = 1$*. Panels (a) and (b) demonstrates fluctuations for an* $N = 64$ *and* $N = 512$ *system respectively. The* $N = 64$ *and* $N = 512$*-body systems have a simulation time of* $t_{sim} = 10^7$ *and* $t_{sim} = 10^9$ *respectively.*

rect sampling: a Nosé-Hoover-Langevin (NHL) method utilises the stochastic noise present in Langevin dynamics [188]. In Figure 1.2, we illustrate typical temperature fluctuations for many-body ($N = 64, 512$) NHL systems. Other reasonable choices of temperature thermostat exist and are discussed in generality by Harish [112]. In Chapter 2, we will be concerned with the NVT (canonical) ensemble. It is worth noting that constant pressure MD (isobaric ensemble) can be achieved with a stochastic Langevin piston [80] and also the Parinello-Rahman [172] method allows for dynamical variation of the simulation cell.

Finally, given the equations of motion, an integrator must be chosen to effectively time step behaviour in the position-momentum phase space. There are many choices for integrators [144] but in Chapter 2, we utilise a simple velocity-Verlet scheme [220] with an NHL thermostat [88]. The velocity-Verlet scheme has the benefit that is a symplectic integrator which is useful for energy conservation purposes [159, 106].

Symplectic maps can be found by splitting/Lie-Trotter methods, but in Chapter 2 we will utilise a NHL time step map (to first order) for N atoms with auxiliary variable ξ, thermostat parameter μ and N_d which denotes the number of spatial degrees of

freedom. The equations of motion are decomposed as follows:

$$d\mathbf{q} = \mathbf{M}^{-1}\mathbf{p}\,dt \tag{1.5a}$$

$$d\mathbf{p} = -\nabla U(\mathbf{q})\,dt - \xi\mathbf{p}\,dt \tag{1.5b}$$

$$d\xi = \mu^{-1}(\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} - N_d k_B T)\,dt - \tilde{\gamma}\xi\,dt - \tilde{\gamma}\mathbf{p}\,dt + \sqrt{2\mu^{-1}\tilde{\gamma}k_B T}\,d\mathbf{W}. \tag{1.5c}$$

Here we can see how the deterministic equations (1.1) can be modelled by Langevin dynamics (1.4) which additionally captures stochastic noise (useful for ergodicity). The addition of a thermostat in the form of an auxiliary variable gives the final equations of motion in (1.5).

The plight of MD is the contention between the computational complexity and the amount of simulation time necessary for accurate sampling/thermodynamic calculation to occur, though recent advancements [141] have alleviated the problem and more topically systems of millions of atoms have been simulated for microsecond time-scales during COVID modelling [11]. Schlick provides a nice review of the advancement of MD [192], in particular pointing to the dawn of machine learning [63].

However, one still aims to sensibly minimise simulation time by either reducing the dimensionality of the system with coarse graining, or by introducing larger time steps. The former is discussed in more detail in Section 1.3, whilst the latter, though of great importance, is not discussed later in this thesis. The choice of time step should be pragmatically chosen [78] such that the numerical integration scheme converges, yet the simulation time is sufficient. As such,multiple time stepping schemes are commonly used when a potential can be decomposed into slow and fast modes [214]. We do not implement such schemes in Chapter 2 though it is worth noting that the time step there is chosen small enough such that the fast dynamics of high temperature systems are well observed.

1.2 Distributions in molecular dynamics

The last subsection has gone some way into explaining how interdisciplinary the field of MD is. As a consequence, much theory has been absorbed from the field of statistical physics where probability distributions are crucial. For MD the main distributions of interest are: ensemble defining, velocity, and force distributions.

We have already mentioned the use of a thermostat to maintain simulations in a given ensemble. There are many such ensembles from classical statistical physics, all characterised by their respective partition functions. The canonical partition function used in Chapter 2 is defined by the Boltzmann distribution which is proportional to $\exp(-\mathcal{H}/k_BT)$, where dynamics are governed by a Hamiltonian $\mathcal{H}$, and k_B is the Boltzmann constant. This distribution underpins Monte Carlo simulations, providing a measure of likelihood that a many-body system is in a particular configuration [170, 171]. Given a probability distribution in configuration space, one is directly able to calculate thermodynamic averages. This is done in Chapter 2 where we calculate the moments of force for a simple liquid of argon [217] in the canonical ensemble.

In the long term, the stochastic equations (1.3) sample the Gibbs-Boltzmann distribution [143]. Langevin dynamics described by equations (1.4) can be infinitely damped to obtain Brownian dynamics [46, 36], the stationary distribution of which is again the Boltzmann distribution [155, 60]. Typically, one can define Fokker-Plank equations used to derive probability distributions arising from SDEs, and it can be shown [50] that a velocity-jump process gives rise to a Maxwell-Boltzmann distribution of velocities, which looks like a weighted Gaussian.

This distribution of velocities is seen ubiquitously in physics, theorised for velocities of idealised fluids in thermodynamic equilibrium [158], Stern [115] experimentally verified this distribution of velocities for a beam of silver atoms in 1920. The Maxwell-Boltzmann distribution of velocities has been used in conjunction with Car-Parrinello MD to explain features of supraconductivity and Bose-Einstein condensates [29]. This

distribution is also important in the random sampling of velocities if one aims to maintain the canonical ensemble [212]. Deviations to the Gaussian form of velocity distributions are often seen in cases where external force fields are present [151, 219].

Systems that have Maxwellian distributions of velocity (in long time) do not necessarily exhibit Gaussian force distributions. This is reported by Shin [198] for an LJ fluid, where deviations from a Gaussian distribution are characterised by a Kullback-Leibler divergence. This has also been reported by Carof [32] for a 3d LJ fluid. In Figure 1.3(a) we illustrate that the second (standardised) moment of the force distribution calculated from MD is not sufficient to accurately represent the observed force distribution, which is addressed in Chapter 2.

Before we detail work on force distributions, we emphasise this is not to be confused with the method of Force Distribution Analysis used in the study of biomolecules whereby the changes of structure in proteins are measured in response to an external force, giving insight into mechanical resistance [203, 204]. Deformations in structure are used to probe the energy landscapes of proteins, aiding the understanding of ligand-receptor binding [123].

Much work has been done in the area of force distributions of many-body systems: with seminal work from Chandrasekhar [33] that employed Markov's theory of random flights to give an expression for the force distribution of a many-body system interacting through a $1/r$ gravitational potential, which was a Holtzmark distribution. Work on gravitational force distributions was extended by Del Popolo [175] to include the rate of change of force. Though we focus on force distributions in the context of MD, stress related contact forces are studied in the field of material science to determine likely conformations of packing arrangements [13, 164].

Force distributions have been widely studied in the area of interacting lattice defects (dislocations), where main consideration is given to nearest neighbour interactions [19]; this is closely connected to the velocity distribution of interacting

(a) (b)

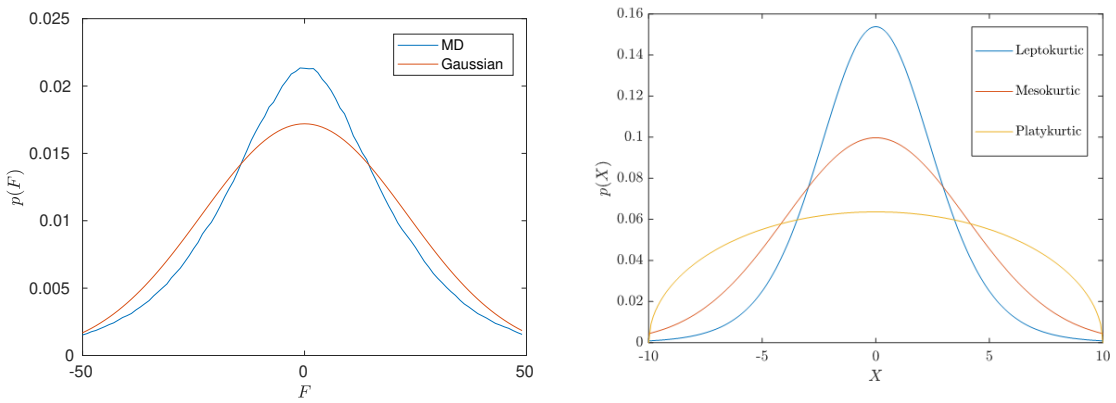


Figure 1.3: *(a) The second moment of force $\langle F^2 \rangle$ is estimated from a simulation (from Chapter 2) of a system of 512 fully atomistic LJ atoms, and used to model a Gaussian distribution with chosen parameters $\sigma^2 = \langle F^2 \rangle$, which is plotted in orange. A histogram method plots the MD results in blue. We see that more parameters are required to model the force distribution than just the second moment of force. (b) We illustrate what a difference in kurtosis looks like for some common probability distributions with a continuous random variable $X \in [-10, 10]$. The blue line represents a leptokurtic distribution (kurtosis greater than 3) with a kurtosis of 4, this is modelled by a Pearson type VII distribution with parameters $m = 5.5$ and $a = 8$. A mesokurtic distribution by definition has a kurtosis of 3, and is modelled here by a Gaussian distribution with $\mu = 0$ and $\sigma = 4$, this is plotted by the orange line. A platykurtic distribution (kurtosis less than 3) with a kurtosis of 2 is plotted by the yellow line; this models a Wigner distribution with $R = 10$.*

vortices in two dimensions for turbulent flow [34]. Force distributions are also found in the study of granular materials [196, 10] and elastic objects where a non-Gaussian distribution characterizes disorder [210].

More recent work has been done with the help of MD by Gabrielli et al [53], who derived an expression for the kurtosis of the force distribution for a lattice system of atoms interacting through the gravitational potential. Further, using density functional theory, an expression for the probability distribution of force for a system interacting through an arbitrary weakly repulsive potential was derived by Rickayzen et al [181, 28].

To characterise the force distribution in Chapter 2, we use the moments from the force distribution. These (even) moments are unbounded on the half line $[0, \infty]$; it is

interesting to note the Stieltjes moment problem. This postulates that only knowing the moments of a distribution on this domain gives rise to the distribution, and if Carleman's condition is satisfied, then it is uniquely so [4]. In principle this means given that the moments of the force distribution satisfy some determinant based condition, one can completely reconstruct the distribution through its moments. As we calculate force statistics 'on the fly' with MD, the proof is not in the scope of this work, and we use moments to address how non-Gaussian a distribution is, though work in Chapter 2 is more general in that it allows us to describe the entire force distribution through understanding all of the moments.

Following work from Erban [49], specific moments of the force distribution estimated from MD simulation have been used to parameterise CG stochastic models for many-body systems as a way to capture this non-Gaussian behaviour. The aim of work in Chapter 2 is to develop the theory detailing how force distributions rely on parameters such as temperature and number density which are easily changed in MD, to aid the development of CG models in the future.

1.3 Force-matching coarse graining

If we understand the moments of velocity and force distributions obtained from all-atom (AA) MD we can better fit coarser models that retain essential features of the higher resolution models [129, 226, 50]. With CG models, we often wish to directly reconcile the energy landscape of the fully atomistic system with a more basic representation maintaining as many physical properties of the system of interest, with as little computational cost as possible [67]. This is illustrated in Figure 1.4 where the energy landscapes from a CG model and AA model are compared.

The review of the field of Coarse Graining is given in more detail than we describe in this section by Noid [166] and more recently by Joshi [129]. However, we will highlight key approaches in the field culminating in force-matching methods. In

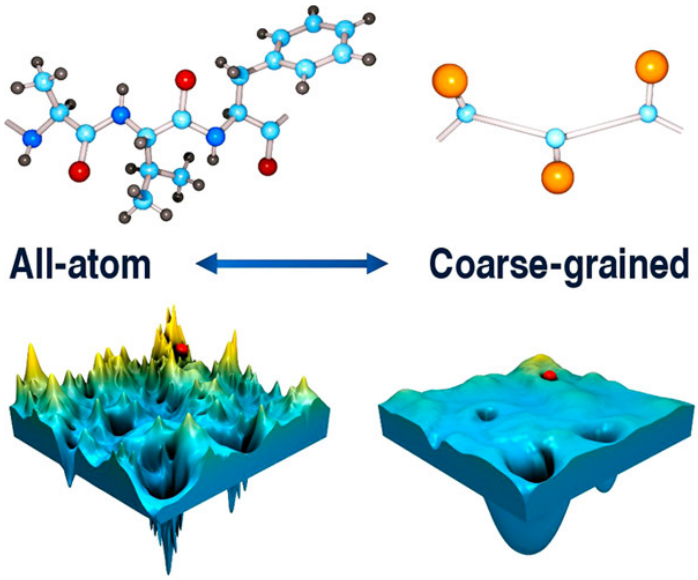


Figure 1.4: *This figure has been reprinted with permission from [77]. This figure illustrates how CG modelling smooths the energy landscape due to a reduction in the degrees of freedom. This is important as the energy minima in the AA simulation can easily trap dynamics in meta-stable states, causing problems with ergodicity as mentioned in Section* 1.1.

essence, CG can reduce the complexity by directly reducing the degrees of freedom by asking: '*what should we consider a CG atom?*'. The superatom approach combines many such atoms into single, larger structures positioned at CG sites. This is commonly used in protein modelling [87], where beads and springs represent collections of thousands of atoms and the bonds connecting them [104, 183], respectively.

Isothermal systems are defined by Helmholtz free energy $F_{\text{helm}} = \langle U \rangle - TS$, where $\langle U \rangle$ is the average internal energy and S is entropy. The probability that a system has a particular energy is proportional to $\exp(F_{\text{helm}}/k_B T)$ due to the fact that $\exp(-F_{\text{helm}}/k_B T)$ is the canonical partition function. The Helmholtz free energy is intrinsically dependent on the configuration $\mathbf{q}$ which is governed by interatomic potential $U(\mathbf{q})$ in the following way:

$$\begin{aligned} \exp(-F_{\text{helm}}/k_B T) &= c \int \exp(-U(\mathbf{q})/k_B T)\, d^3\mathbf{q} \\ &\approx c' \int \exp(-U_{CG}(\tilde{\mathbf{q}})/k_B T)\, d^3\tilde{\mathbf{q}}, \end{aligned} \tag{1.6}$$

where the configuration $\mathbf{q} \in \mathbb{R}^{3N}$ for N atoms is approximated by a reduced configuration of $N_{CG} < N$ larger structures with CG sites given by $\tilde{\mathbf{q}} \in \mathbb{R}^{3N_{CG}}$ which interact via U_{CG}, the CG potential of mean force (PMF). Equation (1.6) is by definition the canonical partition function, where the constant c corresponds to integrating the momentum dependent part of the Boltzmann factor. Maintaining equation (1.6) is of paramount importance for calculating thermodynamic averages [222], though entropy effects are missed and often dynamics are skewed [64, 126]. The reason entropy effects are missed, is because in the procedure of coarse-graining, one reduces the degrees of freedom and essentially encodes the PMF with state dependencies, making U_{CG} a free energy. Approximating U_{CG} the results in an approximation to entropy effects.

One must first define the mapping of the AA configuration to CG sites and then prescribe a force field to this reduced structure model, such that dynamic and thermodynamic behaviour of the AA model is accurately reproduced, for a fraction of the complexity.

Systems can be coarse gained in alternative ways, such as assuming uniform changes in structure (homogeneous deformations) resulting in fewer gradient calculations [186], or by implementing mean field approaches, where single chain field theories have been developed for bio-molecular modelling [55, 42]. This is beyond the scope of the thesis so we will not address these methods again.

Developing force fields for AA MD is expounded in Section 1.6, however there are three main approaches for doing so with a CG system: the top-down approach in which parameterisation of the force field is developed from observed macroscopic behaviour (largely phenomenological), the bottom-up approach which relies heavily on the accuracy of the underlying AA simulation, and the hybrid approach. Transferability, the ability of a CG model to apply to different systems, is the plight of most CG models [74]. However, the top-down approach has been used by the MARTINI force field [81] to avoid the use of reparameterisation for many systems, with con-

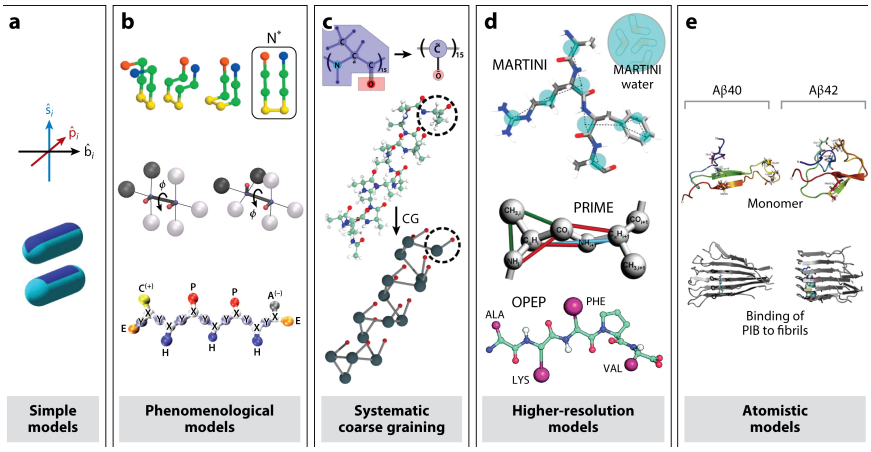


Figure 1.5: *Examples of different resolution models used for protein aggregation are displayed in (a)-(e) from simple models to AA models; it is reprinted with permission from [161]. We see that more degrees of freedom necessarily allow for the models to be described in greater detail, from the orientable stick model in (a) to lattice models in (b) to atomistic models of Alzheimer peptides in (e).*

stants given by a large set of experimental bonding data. A range of CG approaches for a variety of resolutions are illustrated in Figure 1.5.

We focus on bottom-up approaches, as the work in Chapter 2 aids understanding of atomic forces estimated from AA MD, which can be used for future development of CG models. In order to ascertain whether the CG model faithfully reproduces structural, dynamic and thermodynamic features of the AA model, target correlation functions, instead of average thermodynamic quantities, can be used to develop a large class of bottom-up CG methods, following work on reproducibility of correlation functions [86, 35]. We give a short overview of the most commonly used correlation based methods below.

Direct Boltzmann inversion [82] estimates the CG potential directly from atomistic distribution functions. In the case of pair potentials, this method can rely on finding the pair potential of mean force derived from the AA radial distribution function, which can be used to estimate the PMF. This was implemented successfully in RNA modelling [85]. There is no guarantee that the target correlation functions are reproduced, but iterative Boltzmann inversion aims to systematically improve

the CG distributions. It does so by modifying the potential with the difference in CG distribution (obtained from simulation) and target AA distribution [179]. These schemes are popular, though they rely on structural features, and strong correlations between pairs of interactions can lead to issues with convergence [166].

The Inverse Monte Carlo method was introduced by Lyubartsev [153] as an alternative way to systematically improve the CG potential: minimising the difference between the AA and CG distributions. This is done by utilising a susceptibility matrix, each element corresponds to how a random change in PMF affects the overall radial distribution function, which is then used as the basis of a Monte Carlo simulation. This has proved to be a popular scheme [57], though suffers from similar deficiencies as the Boltzmann inversion methods.

Another way to develop CG models is to find an approximate form for the PMF by minimising functionals based on thermodynamics quantities, instead of probability distributions. These are inherently in danger of poor representability, whereby the CG system (and ensuing dynamics) non-accurately describes other thermodynamic variables which are not the target of minimisation [45].

Ercolessi and Adams were the first to propose a force-matching method [51] with application to deriving glue potentials of aluminium, which was mainly used in lattice modelling [103]. Izvekov and Voth extended this idea to form the multi-scale coarse graining (MS-CG) method [125] which provides a systematic way to incorporate force data and iteratively find a many-body CG potential of mean force. This method is more transferable and can be used effectively for modelling mixed bilayers, large proteins, simple liquids, carbohydrates and multi-resolution systems [166].

We highlight the two main components for implementing MS-CG though the reader is referred to [83] for a complete description of this method. One defines a linear mapping from the atomistic Cartesian coordinates $\mathbf{q} = (\mathbf{q}_1, \ldots, \mathbf{q}_N) \in \mathbb{R}^{3N}$

to the CG sites $\tilde{\mathbf{q}} = (\tilde{\mathbf{q}}_1, \ldots, \tilde{\mathbf{q}}_{N_{CG}}) \in \mathbb{R}^{3N_{CG}}$ by:

$$\mathcal{M}_{\tilde{\mathbf{q}}_I}(\mathbf{q}) = \sum_{i=1}^{N} c_{Ii}\mathbf{q}_i = \tilde{\mathbf{q}}_I,$$

for $I = 1, \ldots, N_{CG}$. The collection of these is denoted $\tilde{\mathbf{q}} = \mathcal{M}_{\tilde{\mathbf{q}}}^{N_{CG}}(\mathbf{q})$. In this way the centre of mass of a group of atoms is usually designated as a CG site, where constants c_{Ii} give the appropriate mass dependent weightings.

One can then define the CG PMF $U_{CG}(\tilde{\mathbf{q}})$ in terms of the atomistic potential $U(\mathbf{q})$ with the following equation:

$$\exp(-U_{CG}(\tilde{\mathbf{q}})/k_BT) = c\int \exp\left(-U(\mathbf{q})/k_BT\right)\delta(\mathcal{M}_{\tilde{\mathbf{q}}}^{N_{CG}}(\mathbf{q}) - \tilde{\mathbf{q}})\, d^3\mathbf{q}. \tag{1.7}$$

The CG force field is given by:

$$\mathbf{F}_I(\tilde{\mathbf{q}}) = -\frac{\partial U_{CG}(\tilde{\mathbf{q}})}{\partial \tilde{\mathbf{q}}_I},$$

which can be evaluated in conjunction with (1.7).

A force dependent functional is minimised utilising AA data to derive a potential $U_{CG}(\tilde{\mathbf{q}})$ that accurately matches CG to AA forces. Hence simulation data can be used to form an exact many-body PMF for the CG system. There have been recent developments to MS-CG methods which helps us understand how we can best choose the CG sites [226, 40, 70, 149, 120].

The force-matching procedure estimates (or exactly reproduces) the PMF for the low resolution system defined by CG sites: this gives rise to an energy landscape in which configurations evolve. A large effort has been made to understand how to reproduce these landscapes with force-matching, as this is of paramount importance to maintain vital dynamics of the system [39, 48, 224, 116, 47]. In this way, work

from Chapter 2 can be used to develop bottom-up CG models as we improve the understanding of force distributions. The general concept of force-matching naturally motivates the question: when can we reproduce a configuration that gives rise to known forces?

1.4 Force inversion

The dynamics of a simple liquid is often studied by calculating trajectories in the $6N$-dimensional phase space consisting of the $3N$-dimensional *configuration space* $\mathcal{X} \subset \mathbb{R}^{3N}$ and the $3N$-dimensional momentum space, where N is the number of atoms in the simulated simple liquid. Numerical schemes such as the velocity-Verlet algorithm [220] are used to update the configuration $\mathbf{q} \in \mathcal{X}$ in time. To do this, the interaction potential is used to calculate the force $\mathbf{F}_i$ on every atom, $i = 1, 2, \ldots, N$, by adding the (pairwise) forces between the i^{th} atom and its neighbours

$$\mathbf{F}_i = -\nabla \left(\sum_{j=1, j\neq i}^{N} U\big(r_{ij}\big) \right), \tag{1.8}$$

where r_{ij} is the distance between the i^{th} and j^{th} atoms. Using equation (1.8) for $i = 1, 2, \ldots, N$, we obtain a point $\mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_N) \in \mathcal{F}$ in the corresponding *force space* $\mathcal{F} \subset \mathbb{R}^{3N}$.

The resulting mapping $F : \mathcal{X} \to \mathcal{F}$, defined by $\mathbf{F} = F(\mathbf{q})$, can be obtained using $O(N^2)$ calculations by summing over all pairs of atoms, but it can also be accelerated using a number of methods discussed in Section 1.6.

A force related question of interest, since the 17^{th} century, is the famous N-body problem: after assigning the standard gravitational force field between pairwise interacting bodies, can we solve the dynamics of this chaotic system? Large leaps were made on this problem in the early 20^{th} century by Sundman [209] proving the existence of solutions for the 3-body system. At the turn of the millennium,

Wang [225] generalised these results for $N > 3$. Greengard [102] gives a nice overview of how this problem has percolated through an array of fields in applied mathematics.

The calculation of the inverse mapping

$$F^{-1} : \mathcal{F} \to \mathcal{X} \tag{1.9}$$

is a less studied problem in the literature. This differs from the force inversion problem addressed in the field of atomic force microscopy where displacements as well as forces are known [132]. The problem studied in Chapter 3 relates to the following question: given the instantaneous forces $\mathbf{F} \in \mathcal{F}$, and the pair potential, can we produce a configuration $\mathbf{q} \in \mathcal{X}$ of N atoms that reproduces these forces? In systems that interact classically, such that Newton's equations of motion apply, and via a pairwise potential, such that translational symmetry arises: the configuration $\mathbf{q} \in \mathcal{X}$ which corresponds to given $\mathbf{F} \in \mathcal{F}$ is not unique. However, taking these symmetries into account, the mapping (1.9) is well-defined in the case of harmonic interactions which crucially admit a linear force mapping F. In this case $F(\mathbf{q}) = \underline{\underline{A}}\mathbf{q}$, and $\underline{\underline{A}}$ is a matrix of size $3N \times 3N$ with constant coefficients (demonstrated in Section 3.2.1).

Given force data obtained from AA simulations for an arbitrary potential, one can then use (1.9) to re-position the configuration such that the forces on each tagged atom are exactly the same, but arise from a different potential. In Section 3.2 we show how to go from systems interacting via highly non-linear forces to harmonic forces with force inversion. In this way, any system can be reduced to a purely harmonic system which is less computationally expensive.

For pair potentials like LJ, the mapping $F(\mathbf{q})$ is non-linear and the problem of finding its inverse (*i.e.* reducing any system into a purely LJ system) leads to finding roots of the systems of non-linear equations. The application of iterative methods, like Newton's method, suffers from sensitivity to the choice of initial condition [93]

which leads to an alternative homotopy method developed in Chapter 3 based on incremental deformations from a harmonic system to an LJ system.

1.5 Newton homotopy continuation methods

The first homotopy methods were formulated by Poincaré in the late 19^{th} century as a means to continuously deform topological spaces, which has formed a key topic in the study of algebraic topology. Algebraic topology has pervaded many areas of science, from topological string theory and data analysis to MD [229]. Direct application to MD comes in the form of persistent homology, where the geometric structures of biomolecules can be described, aiding predictions of protein-ligand binding affinities [31]. This is done by taking data points and creating spheres around them, increasing the radii sufficiently will allow overlap and components become connected. Connected components can form or 'birth' holes, and increasing the radii further are the 'death' of these holes. Persistent homology provides understanding of the birth-death cycle of different holes which allows us to understand how the original data points were related.

The application of homotopy methods pertinent to work presented in Chapter 3 relies on the idea of homotopy continuation; in essence this is the tracking of solution paths. Based on work by Hirsch [117], Scarf [190] was the first to develop such tracking schemes to find the fixed points of a square system of nonlinear equations. A detailed overview of path tracking is given by Allgower and Georg [8], in addition to discussions of general piecewise linear continuations and predictor corrector methods.

To motivate the use of homotopy continuation methods, the most ubiquitous scheme used to solve non-linear systems of equations is Newton's method. However, it is well known that this iterative method suffers from sensitivity to initial conditions [233]. Typically, the determinant of the Jacobian (evaluated at the solution) defines how far away from the solution one can make a successful initial guess [93],

which is not known *a priori*.

In order to ameliorate this deficiency, the Newton homotopy continuation method (NHCM) has been developed. It is not to say that this is the only viable approach for the problem; many such developments have come on a more ad-hoc basis, for example methods based on the sign change of the Jacobian have been formulated to deal with tunnel diode problems [27], and this occurred parallel to the path tracking developments in the early 70s. The links between path tracking and Newton's method are outlined by Garcia [94]. Henceforth when we refer to NHCMs, we will be referring to the method developed by Garcia and Zangwill [95] and not those developed by Liao called the homotopy analysis method [197, 1].

Following some notation used by Li [145]: the general premise is that in solving a system of equations defined by $\mathbf{P}(\mathbf{q}) = \mathbf{0} \in \mathbb{R}^N$ for $\mathbf{q} \in \mathbb{R}^N$, we first approach this using a solvable system $\mathbf{Q}(\mathbf{q}) = \mathbf{0} \in \mathbb{R}^N$ and constructing a (linear) homotopy continuation function defined by:

$$\mathbf{H}(\mathbf{q}, \gamma) = (1 - \gamma)\mathbf{Q}(\mathbf{q}) + \gamma\mathbf{P}(\mathbf{q}) = \mathbf{0} \in \mathbb{R}^N. \tag{1.10}$$

In Chapter 3, we will define $\mathbf{Q}(\mathbf{q}) = \mathbf{0}$ to be a the system of equations obtained via purely harmonic interactions, and $\mathbf{P}(\mathbf{q}) = \mathbf{0}$ to arise from purely LJ interactions, so the linear homotopy function gels these two systems together.

For an effective NHCM, we desire three conditions: $\mathbf{Q}(\mathbf{q}) = \mathbf{0}$ is solvable, the set of solutions to $\mathbf{H}(\mathbf{q}, \gamma) = \mathbf{0}$ is finite, and solutions to $\mathbf{P}(\mathbf{q}) = \mathbf{0}$ can be obtained from paths emanating from solutions of $\mathbf{Q}(\mathbf{q}) = \mathbf{0}$. Then solutions to (1.10) can be constructed from $\gamma = 0$ to $\gamma = 1$, thus solving the system $\mathbf{H}(\mathbf{q}, 1) = \mathbf{P}(\mathbf{q}) = \mathbf{0}$. This method is particularly appropriate, as a starting 'guess' is incredibly difficult to formulate for the system $\mathbf{P}(\mathbf{q}) = \mathbf{0}$ arising from LJ atoms, whereas the harmonic case is analytically tractable.

Convergence has been studied with regards to this NHCM in the 1d case by Wu [232], where polynomial and trigonometric equations are considered and are successfully treated with the NHCM in many cases where Newton's method failed. Lee [142] has provided a useful study on the general convergent regions of this convex NHCM form, though computationally determining this is reportedly infeasible for highly non-linear sets of equations. Continuation methods for Newton's method can be tweaked to have global convergence: these have been developed in the field of elastoplastic loading by hybridising with a load-increment method [12].

Sun [208] used this NHCM to investigate the global minimization of the Gibbs free energy in an effort to determine phase equilibria. In the area of geophysical science, resistivity sets are inverted in 1d using a globally convergent NHCM [98]. Recent work in the field of biochemistry has used NHCMs to determine steady steady of chemical reactor models [22]; in mechanical engineering these methods aid investigation of retractable platforms [92].

The use of continuation methods is vast and they are still being developed to improve the speed and radius of convergence for solving. More advanced Newton continuation methods exist: these include the addition of a second auxiliary homotopy function which are used in predictive homotopic path tracking [101], however we use (1.10) in Chapter 3 due to the efficiency of implementation. We implement a novel use of NHCM to aid us in force inversion, which enables us to reconstruct simple pair potential systems (in this case harmonic and LJ) that match target forces exactly.

1.6 Interatomic potentials and non-reciprocal interactions

In Chapters 2-4, we will consider simple liquids described by radially symmetric short-ranged pair potentials (like a LJ liquid), where the $O(N^2)$ scaling of force calculations can also be improved by using a suitable simple interaction cut-off [15], though this can lead to non-physical behaviour [73].

From the computational point of view, the efficient evaluation of $\mathbf{F} = F(\mathbf{q})$ has received a lot of attention in the literature, because the calculation of nonbonded interactions comprise the major bottleneck of classical MD simulations of complex systems [54]. More efficient methods for force evaluation of nonbonded interactions have been proposed, including the Barnes-Hut algorithm [18], Particle Mesh Ewald [37], Fast Multipole Method [182], Multigrid [187] and Multilevel Simulation Methods [223], and the use of Verlet Lists and Linked Cell Methods [7]. Recent advancements in the field have included the implementation of machine learning for force calculations [75].

The theory of classical interatomic potentials has been developed for decades, a review of this research area is provided by Murrell et al [163] or more recently by Ackland [2]. We saw in Section 1.3 a key feature in coarse graining was constructing a potential energy function suitable for the CG system. Many such function choices can naturally arise for a given system [118].

Commonly pair potentials are used to approximate potential energy contributions though caution must be taken to use these appropriately [152] due to neglecting many-body contributions. Despite this: effective pair potentials (see Table 1.1) in many classical circumstances have had fair degrees of success for decades in simulations of liquids [202, 217, 61, 50, 24, 228].

To obtain more accurate results from thermodynamic calculations, many-body contributions are considered in the potential energy function [174, 65]. An example potential incorporating two-body and three-body terms is the Stillinger-Weber potential [207] which accurately incorporates the geometry of silicon, meaning that not only do the pairwise bonds between the silicon atoms matter, but also the triangular sub-structures connecting neighbouring atoms [23]. The embedded atom method potentials [41] incorporate an effective pairwise potential and a density dependent contribution without using the geometric features explicitly.

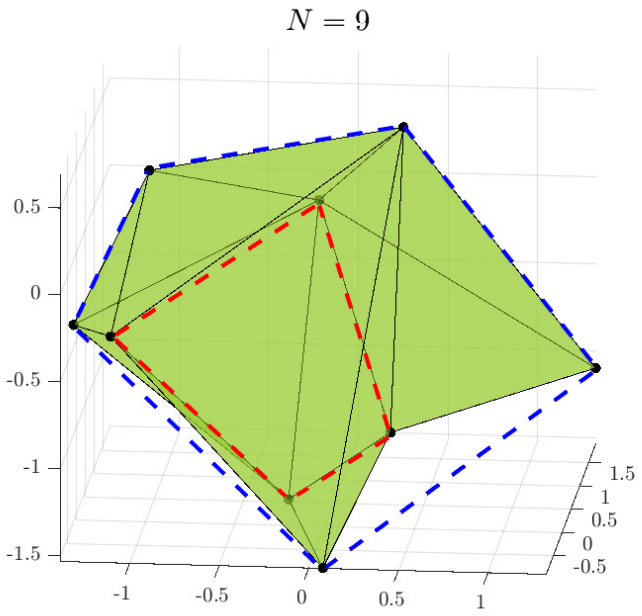


Figure 1.6: *In this figure, we have plotted a* 9 *vertex polyhedron (*$N = 9$*-atom system) highlighting various green triangular faces which would represent* 3*-body interactions, a* 4*-body interaction with the red dashed lines, and a* 5*-body interaction with the blue dashed lines.*

Progressing from pair potentials to those incorporating three-body terms and four-body terms, the most general interatomic potential considered is a sum of all of these contributions, which can also include the one-body terms that arise when an external field is present. The $\tilde{n}$-body terms (for $\tilde{n} \leq N$) are explicitly evaluated given the coordinates of the N atoms: which can be thought of as vertices of a polygon (if co-planar) or a polyhedron. These $\tilde{n}$-body terms in the potential are then thought of as contributions arising from the $\tilde{n}$-polyhedron sub-structures of the shape formed by the vertices, illustrated in Figure 1.6. This forms the basis of fragmentation methods used in ab initio quantum chemistry, a summary and a closed form expression for energy is presented by Richard et al [180].

Tandem to this, cluster descriptions of many-body configurations [189] can also be used in conjunction with $\tilde{n}$-body expansions of the many-body potential [44], this differs from the previous method as it relies on the ordering of vertices as opposed to

their position.

Non-reciprocal interactions are defined as interactions that do not adhere to Newton's third law: that is to say an action does not have an equal and opposite reaction within the confines of pairwise forces [79]. For the case of classical dynamics, forces felt are typically two-body forces and the third law innocuously states $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$, usually potentials depend on interatomic distances r_{ij} [140]. The resulting Lagrangian formed is translationally invariant and the corresponding conserved quantity by Noether's theorem is linear momentum. In a closed system where Noether's theorem holds and where forces are given by two-body interactions, then non-reciprocal interactions imply that linear momentum is no longer conserved.

Seminal examples of non-reciprocal interactions include atomic systems where three-body forces are present (for example three nucleon force in Helium 3 [108], or when the background media moves with the atoms: an example is in relativistic dynamics where linear momentum is not conserved, instead 4-momentum is. This is also seen in electromagnetic interactions when the field moves with the charged particles as well - an example of non-equilibrium physics. A final example of a force that is classically seen to have no reaction is a fictitious force such as the Coriolis force.

Some more recent examples of non-equilibrium physics characterised by non-reciprocal interactions arise in the form of dusty plasma [199, 25], more recently two plasma layers have been formed levitating at different heights, and although the pairwise interactions between dust follow Newton's third laws, the inter-layer forces scale with height and thus impose non-reciprocal forces on each other [59]: these are termed wake mediated interactions. Diffusiophoretic forces found in colloidal physics [62, 201], where a system maintains a concentration gradient in the solute driving non-equilibrium transport due to each colloid's production/consumption of chemicals, can also be used to exhibit non-reciprocal forces. This behaviour is seen

in parallel in the study of micro-swimmers [133, 56] where nanorobots are utilised to probe active transport behaviour.

Robots have been used in the analysis of phase transitions owing to these non-reciprocal interactions [91]; parity breaking organisation of alignment is seen when opposing sets of robots are asked to optimise slightly different objectives. As a final example, water polarization creates an anisotropic medium for interactions to occur, which can affect ion absorption/transport [9].

In Chapter 4, we will predominantly be interested with the analysis of potentials giving rise to reciprocity, with the aim to understand how to produce potentials that give non-reciprocal interactions. A seminal paper from Ivlev [124] overcomes the fact that non-reciprocal forces are non-Hamiltonian and hence cannot be described within the statistical mechanics framework of Hamiltonian dynamics (they violate conditions of variational self adjointness), by creating a pseudo-Hamiltonian which describes the non-equilibrium physics in certain cases. Importantly this analysis is typified by constant non-reciprocity.

Recently there has been an effort to analyse variable non-reciprocity in the context of charged particle interactions and the effect on kinetic energies of a system [148]. Importantly, the result of these interactions is the increase of kinetic energy in the vertical direction. Another consideration for these non-equilibrium systems is the rate of entropy production. It has been shown that non-reciprocal interactions can be dampened sufficiently resulting in equilibrium [237], in cases where they are not, these give rise to non-equilibrium dynamics [150]

A question that has not been answered is characterising when a general inter-atomic potential displays non-reciprocal interactions. It is clear that if the potential depends purely on pairwise distances, then reciprocity is a consequence: so under what symmetries can we conclude that a general position dependent potential function, can be written as function that depends purely on distances? Separately, this is

a fundamental question that underpins classical potential theory, and it is addressed in this thesis in Theorems 1 and 2 in Chapter 4.

Thought into symmetries of a potential has been undertaken by Kinghorn et al [136]: this was used to analyse a specific functional form of potential developed, whilst in Section 4.2 we will be considering a general potential with the goal of understanding when distances are an appropriate variable to describe the potential function. The symmetries required to do this can then be systematically broken to develop models that govern non-reciprocal interactions.

1.7 Structure of thesis

Work undertaken during my DPhil has been partitioned into Chapters 2, 3 and 4 whilst the final Chapter 5 forms a discussion around concluding the work and future directions in scope.

The first novel work is presented in Chapter 2 which is formed from the paper [217] published in Physical Chemistry Chemical Physics, co-authored with Radek Erban. We introduce a statistical mechanics framework to analyse force distributions, in particular the standardised moments of force, and proceed to analyse how these change in asymptotic limits of low temperature and small density. We do this for a simple 1d system and generalise results to an N-body system in 3d, where MD is utilised to compare with predicted behaviour for a system of Argon atoms interacting via a LJ potential, clustering behaviour is seen and compared with literature.

Following this, Chapter 3 is comprised mainly of a second paper in preparation for submission, also co-authored with Radek Erban. We investigate a novel problem that aims to understand how one can go from a force configuration to a position configuration. We begin the work by presenting analytical results for the harmonic potential, and then investigate utilising a NHCM which is the basis for the main body of work in Chapter 3. We work from inverting a harmonic potential to inverting a

purely LJ potential. Algorithms in $N = 2$ dimensions and $N > 2$ dimensions are formalised and investigated.

The final research chapter is presented in Chapter 4 in which we consider a general, position dependent interatomic potential U, and show this can be defined as a function of the interatomic distance variables provided that the potential U satisfies some symmetry assumptions. Moreover, we show that it can be defined as a function of a proper subset of the distance variables, provided that $N > 5$, with the number of distance variables used scaling linearly with N atoms. We then present corollaries and investigate the limitations of implementing a minimal distance description. The theorems and corresponding proofs presented in Chapter 4 have been submitted as a paper [218] written in collaboration with Radek Erban.

Chapter 2

On standardised moments of force distribution in simple liquids

This chapter generalises work submitted for transfer of status by considering an arbitrary standardised moment as opposed to just kurtosis. Presented here is a modified version of my first paper [217]. This chapter differs from the paper in that some of the introduction of the paper has been merged with Chapter 1 of this thesis, and has thus been slightly modified, along with appropriate reformatting changes.

We study the number density and temperature dependence of the force distribution for a many-body system interacting through a Lennard-Jones (LJ) 12-6 potential [128, 230], which is ubiquitously used and has been shown to model homogeneous systems of interacting (Argon) atoms well [177, 220, 109].

In Section 2.2, an in depth investigation is given of the simple two-body system in one spatial dimension, which provides the ideal platform to illustrate the underlying methods while retaining interesting dynamical behaviour. From first principles we derive first-order partial differential equations (PDEs) describing the dependence of the standardised moments of the force distribution has on parameters. In doing so we further derive an analytic expression for the partition function of a two-body

system that depends solely on the standardised moments of the force distribution whereupon the expression is exact in an asymptotic limit of the infinitely dilute system ($n \to 0$). Similarly, an expression is derived relating the average energy of the system to standardised moments of force from the temperature dependent PDE.

In parameter regimes where long-range forces between atoms dominate, we use a truncated Taylor series expansion to derive the leading order behaviour of the kurtosis of the force distribution in the limit $n \to 0$. Finally, we utilise a Laplace integral approximation to ascertain the leading order behaviour of the standardised moments of force at low temperatures ($T \to 0$). Results from simple MD simulation are presented to provide evidence for the efficacy of these methods and underlying assumptions.

This is followed by Section 2.3, where the natural idea that long-range force calculations dictate asymptotic behaviour is extended from the 1d model to many-body systems of arbitrary size in three spatial dimensions. These systems exhibit the physical properties of standard MD simulations: *i.e* cubic geometry with periodic boundary conditions (PBCs) that employ the minimum image convention. This convention means each atom i interacts with the nearest 'image' of atom j, found in surrounding copies of the principle cell, if not the principle cell itself. In this way, we perform calculations on a principal computational cubic cell of a theoretically infinite simple fluid. In Section 2.3.2, MD results are displayed for many-body systems. We present the dependence of the standardised force moments on density, n, and temperature, T, and discuss the parameters and integrator schemes utilised in producing the results of MD simulations

2.1 Notation

Let $\mathbf{F} = (F_1, F_2, F_3)$ denote a force on a tagged atom in a liquid. Considering an isotropic system, the equilibrium distribution of each force coordinate is the same. We

define the standardised moment of the force distribution by using its first coordinate as

$$\alpha_k = \frac{\left\langle F_1^k \right\rangle}{\left\langle F_1^2 \right\rangle^{k/2}}\,, \tag{2.1}$$

where $\left\langle F_1^k \right\rangle$ is the k^{th} moment of the force distribution and α_k standardises the k^{th} moment by scaling it with the k^{th} power of the standard deviation of the force distribution. In a homogeneous fluid, the force distribution will exhibit symmetry around the origin and thus all odd standardised moments vanish, i.e. $0 = \alpha_1 = \alpha_3 = \alpha_5 = \ldots$. As $\alpha_2 \equiv 1$ by definition (2.1), the first non-trivial standardised moment is kurtosis, denoted α_4, which provides a measure of spread that details how tailed the force distribution is relative to a normal distribution [43].

In this chapter, we study how the force distribution depends on the number density of a homogeneous many-body system, and the temperature of the same system in a canonical ensemble. We will do this by studying the behaviour of the second moment of the force distribution $\langle F_1^2 \rangle$ and standardised even moments α_4, α_6, α_8, ... If the force distribution was Gaussian, then the even standardised moments would be

$$\alpha_k = (k-1)!! = \prod_{i=1}^{k/2} (2i-1), \quad \text{for } k = 2, 4, 6, 8, 10, \ldots, \tag{2.2}$$

and the second moment $\langle F_1^2 \rangle$ would be sufficient to parametrize the force distribution. However, the force distributions in simple liquids have been reported to deviate from Gaussian distribution [198, 49, 32]. Thus our analysis will also tell us how non-Gaussian the force distribution is.

We consider a system of N identical atoms interacting via the LJ 12-6 potential [128]. This is a ubiquitous interatomic pairwise potential; here the potential between atoms labelled $i, j = 1, 2, \ldots, N$ positioned at $\mathbf{q}_i, \mathbf{q}_j \in \mathbb{R}^3$ is given (in re-

duced units [90]) by the expression

$$U_{ij}(r_{ij}) = 4\left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^{6}}\right), \tag{2.3}$$

where $r_{ij} = |\mathbf{q}_i - \mathbf{q}_j|$ is the distance between atoms. The LJ potential (3.25) between two atoms has a unique minimum obtained at $r_{ij} = r^* = 2^{1/6}$.

We employ the framework of statistical mechanics for this closed many-body system and describe atom $i = 1, 2, \ldots, N$ by phase space coordinates $\{\mathbf{q}_i, \mathbf{p}_i\} \in \mathbb{R}^6$, were $\mathbf{p}_i$ denotes the momentum of the i^{th} atom. We work in the canonical ensemble with temperature T; the partition function therefore becomes

$$\mathcal{Z}_N(T, V) = \frac{1}{h^{3N}\, N!} \iint\limits_{\Omega_\mathbf{q} \times \Omega_\mathbf{p}} \exp[-\beta \mathcal{H}(\mathbf{q}, \mathbf{p})]\, \mathrm{d}^3\mathbf{q}\, \mathrm{d}^3\mathbf{p}\,, \tag{2.4}$$

where V is the volume of our closed system, and $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N)^T$ and $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N)^T$ are vectors containing the positions and momenta of all atoms in the system. Our integration domain is given by $\Omega_\mathbf{q} \times \Omega_\mathbf{p} \subset \mathbb{R}^{3N} \times \mathbb{R}^{3N}$. This denotes the phase space of our system. For systems of interest $\Omega_\mathbf{p} \equiv \mathbb{R}^{3N}$. The underlying geometry of the system (and principle simulation cell) is a cubic box of size $L > 0$, therefore $\Omega_\mathbf{q} \equiv (-L/2, L/2] \times \cdots \times (-L/2, L/2]$. The phase space volume elements in equation (2.4) are denoted by

$$\mathrm{d}^3\mathbf{q} = \prod_{i=1}^{N} \mathrm{d}^3\mathbf{q}_i \qquad \text{and} \qquad \mathrm{d}^3\mathbf{p} = \prod_{i=1}^{N} \mathrm{d}^3\mathbf{p}_i. \tag{2.5}$$

Throughout this work we make use of reduced units [90], utilising Argon parameters [185]. In particular, all instances of T in this work can be translated back to SI units with the transformation $T \to k_B T$ where k_B is the Boltzmann factor. Therefore, in the partition function (2.4), we have $\beta = 1/T$ and h is the Planck constant (≈ 2.17 in reduced units). Finally, $\mathcal{H}(\mathbf{q}, \mathbf{p})$ is the classical Hamiltonian

$\mathcal{H}(\mathbf{q},\mathbf{p}) = K(\mathbf{p}) + U(\mathbf{q})$ with kinetic energy $K(\mathbf{p}) = |\mathbf{p}|^2/2$ (where the usual factor of mass is unity under reduced units) and a general potential $U(\mathbf{q})$.

The statistical average of a quantity X for this N-body system is given by

$$\langle X \rangle = \frac{1}{\mathcal{Z}_N \, h^{3N} \, N!} \iint\limits_{\Omega_{\mathbf{q}} \times \Omega_{\mathbf{p}}} X \, \exp[-\beta \mathcal{H}(\mathbf{q},\mathbf{p})] \, \mathrm{d}^3\mathbf{q} \, \mathrm{d}^3\mathbf{p} \,, \tag{2.6}$$

where the Boltzmann factor acts as a statistical weighting for a configuration $\{\mathbf{q},\mathbf{p}\} \in \mathbb{R}^{6N}$, normalised such that $\langle 1 \rangle = 1$.

We label atoms so that the first one is the tagged atom. Denoting the force on the tagged atom produced from the j^{th} atom by $\mathbf{F}_j = (F_{j,1}, F_{j,2}, F_{j,3}) \in \mathbb{R}^3$, for $j = 2, 3, \ldots, N$, the total force $\mathbf{F} = (F_1, F_2, F_3)$ on the tagged atom is

$$\mathbf{F} = \sum_{j=2}^{N} \mathbf{F}_j .$$

We define

$$f_k = \int_{\Omega_{\mathbf{q}}} \left(\sum_{j=2}^{N} F_{j,1} \right)^k \exp[-\beta \, U(\mathbf{q})] \, \mathrm{d}^3\mathbf{q} \tag{2.7}$$

for $k = 0, 1, 2, \ldots$. Then we have

$$\frac{f_k}{f_0} = \left\langle \left(\sum_{j=2}^{N} F_{j,1} \right)^k \right\rangle = \langle F_1^k \rangle .$$

Then the k^{th} standardised moment (2.1) is given by

$$\alpha_k = \frac{f_0^{k/2-1} f_k}{f_2^{k/2}} \,, \tag{2.8}$$

where we are interested in cases $k = 4, 6, 8, \ldots$.

In order to study how the force distribution depends on the physical parameters of interest it is useful to identify how changes in these parameters will manifest

themselves in the system. Indeed, we choose to work in the canonical ensemble with a target temperature of T: this is accomplished with the use of a thermostat which is discussed further in Section 2.3.2 and Appendix 2.6. It is more illuminating to see that if we have a system with a fixed number of free interacting atoms N in a cubic box of side L; the (reduced) number density is given by $n = N/L^3$. Therefore the approach we employ in this chapter to ascertain how values of standardised moments depend on number density, will be to keep the number of atoms fixed but vary the box width L - this will manifest as a change in density n. Similarly one could keep the volume of the cubic box the same and vary the number of atoms though this is a point of discussion in Section 2.3.2.

For the remainder of the chapter we will study systems of varying dimensions and size, using equation (2.8) as a crucial starting point in each calculation. We will naturally proceed by investigating systems of increasing complexity; starting from a cartoon one-dimensional model culminating to a general many-body system of arbitrary size.

2.2 One atom in a potential well

We now go on to illustrate three approaches to obtain the dependence of the force distribution on parameters n and T. It is useful to note that, as we are now working in one spatial dimension, density n is proportional to $1/L$, *i.e.* we have $n \propto 1/L$.

We will consider a simple system in one spatial dimension consisting of two atoms interacting through the LJ potential (3.25) in interval $[0, L]$ with PBCs. One of the atoms is considered to be fixed at position $q_0 = L/2 \in [0, L]$ and the other atom is free to move, therefore, we have $N = 1$ free atom. Its position is denoted $x \in [0, L]$. Therefore, the interatomic distance is $r = |x - q_0|$. Using our simplified

one-dimensional set up, $F_1 = F$ and $\Omega_{\mathbf{q}} = (0, L)$, equation (2.7) reduces to

$$f_k(L) = \int_0^L F^k(|x - q_0|)\, \exp[-\beta\, U(|x - q_0|)]\, \mathrm{d}x, \tag{2.9}$$

which is the marginalised expected value of the k^{th} moment of force $F(x) = -\mathrm{d}U/\mathrm{d}x$, where we have dropped subscripts in the LJ potential (3.25) and we write it as $U(z) = 4(z^{-12} - z^{-6})$.

Utilising the symmetry of the potential (and therefore the force) we are left with

$$f_k(L) = 2 \int_0^{L/2} F^k(r)\, \exp[-\beta\, U(r)]\, \mathrm{d}r. \tag{2.10}$$

In what follows, we will assume that we are in a regime where the box width L satisfies $L \gg r^*$, where $r^* = 2^{1/6}$ minimizes the LJ potential U.

2.2.1 Differential equation for standardised moments

We consider a perturbation of the form $L \to L + \delta L$. Using equation (2.10) and considering terms to the order $O(\delta L)$, we obtain

$$\begin{aligned} f_k(L + \delta L) &= f_k(L) \,+\, f_k'(L)\, \delta L \,+\, O\big(\delta L^2\big) \\ &= f_k(L) \,+\, F^k(L/2)\, \exp[-\beta\, U(L/2)]\, \delta L \,+\, O\big(\delta L^2\big). \end{aligned}$$

Using equation (2.8), we approximate $\alpha_k(L + \delta L)$ by

$$\alpha_k(L) + \alpha_k(L)\, \upsilon_k(L)\, \exp[-\beta\, U(L/2)]\, \delta L + O\big(\delta L^2\big),$$

where our notation $\alpha_k(L)$ highlights the dependence of the standardised moments of force, α_k, on L, and function $\upsilon_k(L)$ is given by

$$\upsilon_k(L) = \frac{k-2}{2\,f_0(L)} + \frac{F^k(L/2)}{f_k(L)} - \frac{k\,F^2(L/2)}{2\,f_2(L)}\,. \tag{2.11}$$

Taking the limit $\delta L \to 0$, we obtain the derivative of the k^{th} standardised moment of force, with respect to L, as

$$\frac{\partial \alpha_k}{\partial L}(L) = \upsilon_k(L)\,\exp[-\beta\,U(L/2)]\,\alpha_k(L), \tag{2.12}$$

where $\upsilon_k(L)$ are expressed in terms of integrals (2.10) as given by equation (2.11).

2.2.2 Far-field integral approximation

To further analyze integrals (2.10), we introduce a cut-off c, which satisfies that $r^* < c < L/2$, where $r^* = 2^{1/6}$ is a unique maximum of $\exp[-\beta\,U(z)]$, which can be Taylor expanded as $1 + 4\beta z^{-6} + \ldots$. Considering sufficiently large L, we can choose the cut-off c, so that

$$\left| f_0(L) - 2\left(\int_0^c \exp[-\beta\,U(r)]\,\mathrm{d}r + \int_c^{L/2} 1 + \frac{4\beta}{r^6}\,\mathrm{d}r \right) \right| \leq \varepsilon, \tag{2.13}$$

where tolerance ε is chosen to be 10^{-4} in our illustrative computations. This splitting allows us to numerically calculate the bulk of the integral (2.10) as a constant independent of L and then use the second term to give an analytic expression for α_k with dependence on L, and ultimately on n.

The range of values of T that are of typical use are chosen in order to maintain the liquid state of Argon during simulation. These are approximately temperatures in the interval $0.70 < T < 0.73$ under ambient conditions [147]. Therefore, as volume is varied we are in a regime where $\beta = o(1)$, for convenience we set $\beta = 1$. Though

given that the density of our system changes between each simulation some systems will be in a liquid phase and others in a gaseous phase, this is a point of discussion in Section 2.3.2.

Splitting the integration domain $[0, L/2]$ of integral (2.10) into $[0, c]$ and $[c, L/2]$, we use the exact form of the integrand in $[0, c]$ to obtain a 'near-field' contribution. Utilising an approximate form for the integrand given by the truncated Taylor expansion $f(z)$ in the domain $[c, L/2]$ gives rise to a density dependent 'far-field' contribution. Combining these we arrive at the approximate form for $f_0(L)$. Using cut-off $c = 2$, equation (2.13) is satisfied with $\varepsilon = 10^{-4}$. Therefore, upon numerically calculating the bulk contribution for the integral with domain $[0, 2]$, we get

$$f_0(L) = 2 \int_0^{L/2} \exp[-\beta\, U(r)]\, \mathrm{d}r \;=\; b_0 + L + o\big(L^{-4}\big) \tag{2.14}$$

with $b_0 = -0.71832$, which depends on our choice of cut-off $c = 2$. Similarly, we can calculate far-field integral approximations of integrals (2.10) for general values of $k = 2, 4, 6, 8, 10, 12$. The integrand $F^k(r) \exp[-\beta\, U(r)]$ has maxima when $r = r^* = 2^{1/6}$ or when $k\, U''(r) = \beta\, (U'(r))^2$. This forms a cubic in r^6 that can be solved. For the values of k used in this work, this sometimes results in a global maximum, that always lies at a distance less than $r < r^*$ from the origin. Therefore $r^* = 2^{1/6}$ is the furthest maximum of the integrand from the origin.

Splitting integral (2.10) into a near-field and far-field contribution, using the general cut-off $c = 2$, we find

$$f_k(L) \;=\; b_k + o\big(L^{-7k+2}\big) \tag{2.15}$$

The near-field contributions, b_k, generally increase vastly if we increase the value of k,

for example

$$b_0 = -0.71832, \;\; b_2 = 130.64 \;\text{ and }\; b_4 = 2.5727 \times 10^5, \tag{2.16}$$

while the dependence on L decreases more rapidly for larger values of k. Therefore, the non-negligible density contributions to $\alpha_k(L)$ in the low density limit come exclusively from the normalisation $f_0(L)$ given by (2.14).

Substituting equations (2.14) and (2.15) in equation (2.8), we obtain an expression for the general k^{th} standardised moment of force

$$\alpha_k(L) = \frac{b_0^{k/2-1} b_k}{b_2^{k/2}} \left(1 + \frac{L}{b_0} + o\big(L^{-4}\big)\right)^{k/2-1}. \tag{2.17}$$

Using the values of b_0, b_2 and b_4 given by (2.16), we obtain the dependence of the kurtosis of the force distribution on the reduced number density $n = 1/L$ in the dilute limit $n \to 0$ as $\alpha_4 = -10.828 + 15.074\, n^{-1} + o(n^4)$. Figure 2.1 compares this result with the results obtained by MD simulation of the one atom system. We observe that MD is in good agreement with the results obtained by formula (2.17).

2.2.3 Leading order behaviour for differential equation (2.12)

Since $L/2 > r^*$, the force $F(L/2)$ monotonically decreases as a function of L. When looking at leading order approximations in the low density limit $n \to 0$ (equivalent to limit $L \to \infty$) to equation (2.12), we need to analyse $\upsilon_k(L)$. The second and third term in equation (2.11) converge to zero more rapidly than the first term as $L \to \infty$, therefore the leading order behaviour is given by the first term

$$\upsilon_k(L) \sim \frac{k-2}{2\, f_0(L)} \qquad \text{as} \qquad L \to \infty. \tag{2.18}$$

By utilising the far-field integral approximation (2.14), we arrive at $f_0(L) \sim (b_0 +$

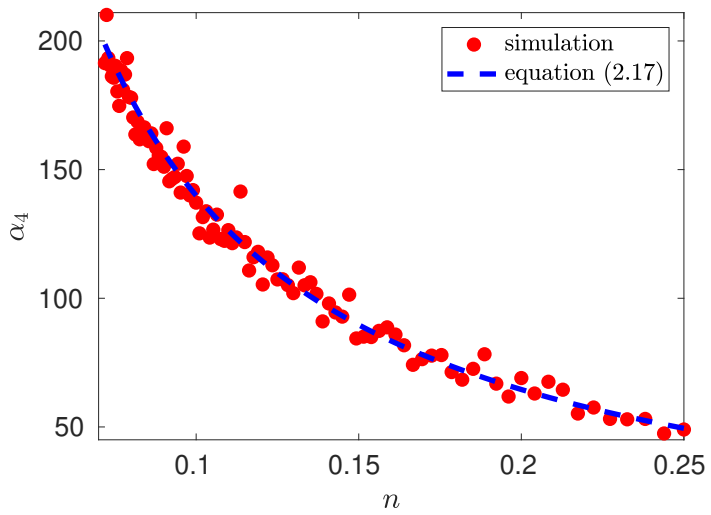


Figure 2.1: *Plot of α_4 as a function of $n = 1/L$ for the illustrative one-atom system. Results of MD simulations are compared with $\alpha_4 = -10.828 + 15.074\, n^{-1}$ obtained by using equation* (2.17) *with b_0, b_2 and b_4 given by* (2.16) *(blue dashed line). MD simulation results utilising Langevin dynamics [100] described in equation* (2.44)*, with friction parameter $\tilde{\gamma} = 0.1$ are represented by red dots. The MD simulation length was a total of 1.1×10^8 time steps with the first 10^7 time steps used for initialisation.*

L), where $b_0 = b_0(c)$ is a constant term that depends on cut-off parameter c. With this, our leading order approximation of the k^{th} standardised moment, α_k^0, obeys

$$\frac{\partial \alpha_k^0}{\partial L}(L) = \frac{k-2}{2\,(b_0 + L)}\,\alpha_k^0(L)\,.$$

Finally this gives us that

$$\alpha_k^0(L) = C_k\,(b_0 + L)^{k/2-1} = C_k\left(b_0 + n^{-1}\right)^{k/2-1}, \tag{2.19}$$

where $n = 1/L$ is the reduced number density and C_k is a constant. Equation (2.19) gives the same leading order behaviour $n^{1-k/2}$ in the limit $n \to 0$ as equation (2.17): the same behaviour is also seen for the LJ fluid in Section 2.3. Though the method above is more generally applicable to include potentials that monotonically decay as r^{-a} as $r \to \infty$ for $a > 0$.

We next make the observation that equation (2.4) in 1d can be written as:

$$\mathcal{Z}_1(T,V) = \frac{1}{h}\int_0^L \exp[-\beta\, U(q)]\ \mathrm{d}q \int_{-\infty}^{\infty} \exp\left[-\frac{\beta\, p^2}{2}\right]\ \mathrm{d}p\,, \tag{2.20}$$

where the Planck factor of $1/h$ arises instead of $1/h^3$ due to the fact that we are in one-dimensional physical space. Using (2.10), we obtain

$$f_0(L) = h\, \mathcal{Z}_1(T,V)\, \sqrt{\frac{\beta}{2\pi}}. \tag{2.21}$$

Considering the low density limit $n \to 0$ (i.e. $L \to \infty$) in equation (2.12) and using (2.18) and (2.21), we obtain

$$\mathcal{Z}_1(T,V) \sim \frac{(k-2)\sqrt{\pi}}{\sqrt{2h^2\,\beta}}\ \left[\alpha_k(L)\left(\frac{\partial \alpha_k}{\partial L}(L)\right)^{-1}\right], \tag{2.22}$$

as $L \to \infty$, which is true for each $k = 4, 6, 8, \ldots$. In particular, we can obtain the partition function (2.20) in the dilute (low density) limit by using information about the moments of the force distribution.

The accuracy of equation (2.22) is illustrated in Figure 2.2, where we use $k = 4$. We use MD simulations of a single atom, using a range of simulation box widths L. We estimate the values of kurtosis of the force distribution, its derivative with respect of L and use the right hand side of equation (2.22) to estimate the $\mathcal{Z}_1(T,V)$. Considering $L \geq 10$, the result is within 5% error when compared with the exact result (2.20), while for larger values of box width L the error decreases to around 1%, confirming that the formula (2.22) is valid in the asymptotic limit $L \to \infty$.

2.2.4 Temperature dependence of standardised moments

One can perform a similar analysis as in Section 2.2.1, viewing the moments $\alpha_k = \alpha_k(T)$ as a function of temperature $T = 1/\beta$. To do that, we consider the moment

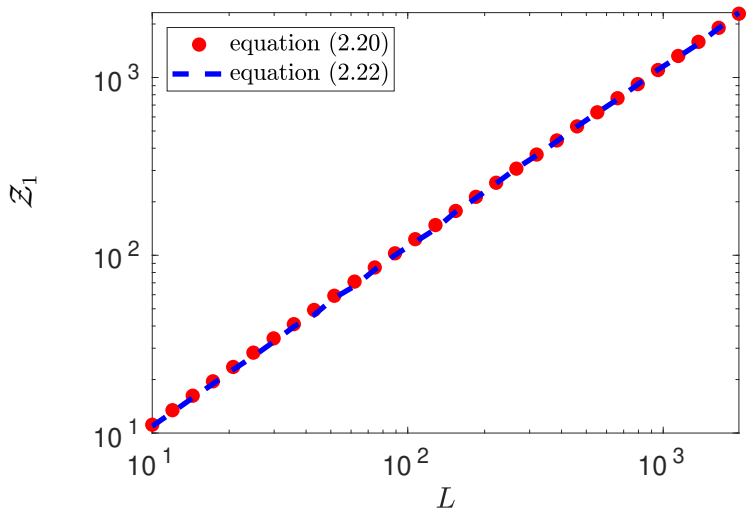


Figure 2.2: *Approximation of the partition function $\mathcal{Z}_1(T, V)$ obtained using the right hand side of equation* (2.22) *with $k = 4$ and values of kurtosis (α_4) estimated from MD simulation (blue dashed line). The exact values obtained by* (2.20) *are plotted as the red dots.*

definition (2.10) as a function of temperature T, namely, we define

$$f_k(T) = 2 \int_0^{L/2} F^k(r) \, \exp\left[-\frac{U(r)}{T}\right] \, \mathrm{d}r. \tag{2.23}$$

Considering small perturbations of these functions with respect to $T \to T + \delta T$, while fixing the domain length L, and collecting terms up to first order in δT, we obtain

$$\frac{\partial \alpha_k}{\partial T}(T) = \nu_k(T) \, \alpha_k(T) \, , \tag{2.24}$$

where

$$\nu_k(T) = \left(\frac{k}{2} - 1\right) \frac{f_0'(T)}{f_0(T)} + \frac{f_k'(T)}{f_k(T)} - \left(\frac{k}{2}\right) \frac{f_2'(T)}{f_2(T)} \, . \tag{2.25}$$

Combining equations (2.24) and (2.25) with equation (2.21) where $\beta = 1/T$, we obtain

$$\frac{\partial}{\partial T} \log\left(\frac{\alpha_k^2 f_2^k}{f_k^2}\right) = (k-2)\left(\frac{\partial}{\partial T} \ln(\mathcal{Z}_1) - \frac{1}{2T}\right).$$

Since $-\partial/\partial\beta(\ln \mathcal{Z}_1)$ is equal to the average energy of the system, $\langle E\rangle$, we have for each $k = 4, 6, 8, \ldots$

$$\langle E\rangle = \frac{T}{2} + \frac{T^2}{k-2}\frac{\partial}{\partial T}\log\left(\frac{\alpha_k^2 f_2^k}{f_k^2}\right), \tag{2.26}$$

where the first term on the right hand side of equation (2.26) is the average kinetic energy of our one-atom system. Substituting equation (2.8) into the second term on the right hand side, it can be rewritten as $T^2\partial(\ln f_0)/\partial T$. Thus, using equation (2.6), we confirm that the second term on the right hand side of equation (2.26) is the average potential energy as

$$T^2\partial(\ln f_0)/\partial T = \frac{\int_0^{L/2} U(r)\exp\left[-\frac{U(r)}{T}\right]\,\mathrm{d}r}{\int_0^{L/2}\exp\left[-\frac{U(r)}{T}\right]\,\mathrm{d}r} = \langle U\rangle.$$

2.2.5 Low temperature limit

Next, we consider the behaviour of the k^{th} standardised moment of force, $\alpha_k(T)$, given by equation (2.8), in the low temperature limit, $T \to 0$, which is equivalent to the limit $\beta \to \infty$. Since the interatomic potential $U(r)$ has a global minimum at $r = r^*$ in interval $[0, L/2]$, integrals of the form (2.10) and (2.23) can be approximated by Laplace's method in the limit $\beta \to \infty$ and $T \to 0$, respectively. A general discussion of Laplace's method is given in Chapter 6 of the book by Bender and Orszag [20].

We calculate the asymptotic expansion of $f_0(T)$ by applying Laplace's method to integral (2.23) for $k = 0$. We approximate the integration limits of integral (2.23) to lie within the domain $r \in (r^* - \varepsilon, r^* + \varepsilon)$, where $\varepsilon \ll 1$, and we Taylor expand $U(r)$ at $r = r^*$. Using $U'(r^*) = 0$, we have

$$\begin{aligned} U(r) \approx\ & U(r^*) + (r - r^*)^2 U''(r^*)/2 \\ & + (r - r^*)^3 U^{(3)}(r^*)/6 + (r - r^*)^4 U^{(4)}(r^*)/24\,, \end{aligned}$$

where we denote the m^{th} derivative of U as $U^{(m)}$ for $m \geq 3$. Substituting into integral (2.23), we arrive at the asymptotic expansion

$$f_0(T) \sim \frac{\sqrt{\pi\, T}\, \exp[-U(r^*)/T]}{\sqrt{2\, U''(r^*)}} \Big[1 + A_1\, T + o(T)\Big], \tag{2.27}$$

as $T \to 0$, where constant A_1 is given by [20]

$$A_1 = \frac{5\,(U^{(3)}(r^*))^2}{24\,(U''(r^*))^3} - \frac{U^{(4)}(r^*)}{8\,(U''(r^*))^2}\,.$$

To apply Laplace's method to integral (2.23) for $k = 2, 4, 6, \ldots$, we note that $F^k(r) = (U'(r))^k$ for even values of k. Using the truncated Taylor expansion around $r = r^*$ and noting $U'(r^*) = 0$, we have

$$F^k(r) \approx (r - r^*)^k \left((U''(r^*))^k + (r - r^*)\, C_{k,1} + (r - r^*)^2\, C_{k,2} \right),$$

where $C_{k,1}$ and $C_{k,2}$ are constants, which can be expressed (see Appendix 2.5) in terms of the derivatives of potential $U(r)$ at $r = r^*$ as

$$C_{k,1} = \frac{k\,(U''(r^*))^{k-1} U^{(3)}(r^*)}{2} \tag{2.28}$$

$$C_{k,2} = \frac{k}{24}\,(U''(r^*))^{k-2} \times \left(3(k-1)\left(U^{(3)}(r^*)\right)^2 + 4\, U''(r^*)\, U^{(4)}(r^*) \right). \tag{2.29}$$

This gives the asymptotic expansion

$$f_k(T) \sim \frac{\sqrt{\pi\, T}\, \exp[-U(r^*)/T]}{\sqrt{2\, U''(r^*)}} \times \left[B_{k,1}\, T^{k/2} + B_{k,2}\, T^{k/2+1} + o\left(T^{k/2+1}\right) \right]. \tag{2.30}$$

as $T \to 0$, where constants $B_{k,1}$ and $B_{k,2}$ are given by

$$B_{k,1} = \left(U''(r^*)\right)^{k/2} (k-1)!!$$

and

$$B_{k,2} = \frac{C_{k,2}(k+1)!!}{|U''(r^*)|^{k/2+1}} - \frac{U^{(3)}(r^*)\, C_{k,1}(k+3)!!}{6\,|U''(r^*)|^{k/2+2}} + \frac{(U''(r^*))^{k/2-2}(k+3)!!}{24}\left(\frac{(k+5)\,U^{(3)}(r^*)}{3\,U''(r^*)} - U^{(4)}(r^*)\right).$$

Substituting (2.27)-(2.30) into (2.8) gives the expression as $T \to 0$:

$$\alpha_k \sim \frac{B_{k,1}}{B_{2,1}^{k/2}}\left(1 + T\left[\left(\frac{k}{2} - 1\right)A_1 + \frac{B_{k,2}}{B_{k,1}} - \frac{k\,B_{2,2}}{2\,B_{2,1}}\right]\right). \tag{2.31}$$

In particular, we have

$$\begin{aligned} \alpha_2 &\sim 1 + O(T^2), \\ \alpha_4 &\sim 3.000 + 45.084\,T + O(T^2). \end{aligned} \tag{2.32}$$

Therefore, Laplace's method predicts that the kurtosis, $\alpha_4(T)$, tends to 3 in the low temperature limit, and to leading order the perturbation away from this value scales linearly in temperature. This limiting behaviour is to be expected as during the Laplace approximation we use a Gaussian distribution to approximate the Boltzmann factor, and the kurtosis of any Gaussian distribution is identically equal to 3. We can interpret this approach as approximating the force distribution as Gaussian and perturbations of the system around small temperatures give rise to non-Gaussian contributions to the standardised moments.

Results from MD simulation are illustrated in Figure 2.3. Due to the rapid nature of the growth of F^k, one has to restrict to a smaller range of temperature to exhibit the linear growth predicted. This holds for the range of $k = 2, 4, 6, 8, 10, 12$

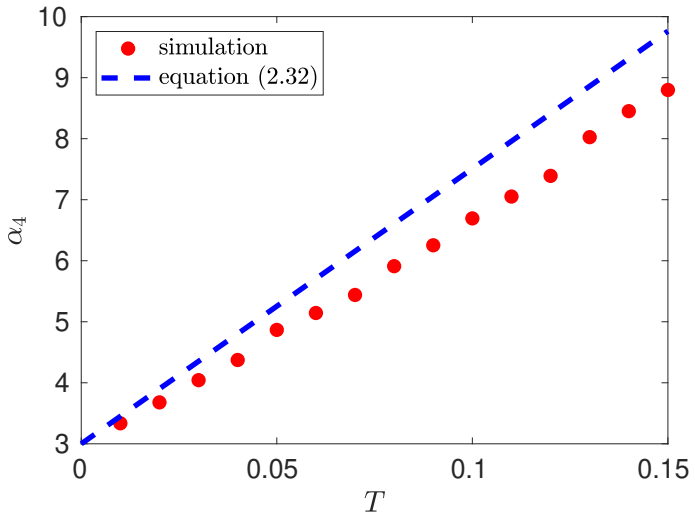


Figure 2.3: *Kurtosis, α_4, as a function of temperature, T, for $T \leq 0.15$. The linear behaviour is estimated as $\alpha_4(T) \sim 2.9388 + 37.002\,T$ for $T \in (0.01, 0.10)$ (using the MD computed data visualized as red dots). We compare this to the theoretical result predicted by equation* (2.32) *(illustrated by the blue dashed line).*

tested in this work. This can be explained by the fact that the integrand of equation (2.23) sometimes admits 2 maxima in the range $(0, \infty)$. The global maximum of the integrand does not always correspond with the global minimum of the potential $U(r)$ and therefore the global maximum of $\exp[-\beta U(r)]$. We therefore need to probe smaller temperature ranges to more accurately match the predicted behaviour in equation (2.32), though maintaining the canonical ensemble at such minute temperature ranges necessitates coupling the simulation thermostat tightly, which may affect dynamics [137].

2.3 Many-body systems

In this section we employ the far-field approximation approach introduced in Section 2.2.2 and we will vary the number density of the system by changing the size L of the integration domain, which will be given as the three-dimensional cube $[0, L]^3$. Using notation introduced in Section 2.1, the distance between atoms labelled $i, j = 1, 2, \ldots, N$ positioned at $\mathbf{q}_i, \mathbf{q}_j \in \mathbb{R}^3$ is denoted by $r_{ij} = |\mathbf{q}_i - \mathbf{q}_j|$. Taking into

account the PBCs, the distance $|\mathbf{q}_i - \mathbf{q}_j|$ is the minimum image interatomic distance given by

$$|\mathbf{q}_i - \mathbf{q}_j| = \left(\overline{(q_i^x - q_j^x)}^2 + \overline{(q_i^y - q_j^y)}^2 + \overline{(q_i^z - q_j^z)}^2 \right)^{1/2}, \tag{2.33}$$

where the overline denotes $\overline{\zeta} = \zeta - L \lfloor \zeta / L \rceil$ for $\zeta \in \mathbb{R}$ and $\lfloor \cdot \rceil$ rounds a real number to the nearest integer. For an interacting N-body system the dimensionality of the integral given by equation (2.7) is $3N$. We first present an illustrative calculation with $N = 2$ interacting atoms in Section 2.3.1 and then we study systems with larger values of N in Section 2.3.2.

2.3.1 Dependence of α_k on density for $N = 2$ interacting atoms

In Section 2.2, we have considered two atoms in the one-dimensional spatial domain, where one atom was fixed at position q_0, i.e. we have effectively studied a single atom in a one-dimensional potential well. Here, we will consider $N = 2$ interacting atoms in the three-dimensional cubic domain $[0, L]^3$ with PBCs. We calculate the k^{th} standardised moment of force according to equation (2.8). To do so, we consider equation (2.7), where we have $\mathrm{d}^3\mathbf{q} = \mathrm{d}^3\mathbf{q}_1 \, \mathrm{d}^3\mathbf{q}_2$, $U(\mathbf{q}) = U(r_{12})$, $F_1(\mathbf{q}) = F_1(r_{12})$ and we integrate over the domain $\Omega = [0, L]^3 \times [0, L]^3$ to get

$$f_k = \int_\Omega F_1^k(r_{12}) \, \exp[-\beta \, U(r_{12})] \, \mathrm{d}^3\mathbf{q}_1 \, \mathrm{d}^3\mathbf{q}_2. \tag{2.34}$$

It is useful to introduce a change of coordinates $\xi^\ell = q_1^\ell - q_2^\ell$ and $\eta^\ell = q_1^\ell + q_2^\ell$ for $\ell = x, y, z$. We note that r_{12} is only dependent on the ξ^ℓ variables, therefore one can trivially integrate (2.34) through the η^ℓ variables as the integrand has no dependence on these to obtain

$$f_k = \frac{L^3}{8} \int_{-L}^{L} \int_{-L}^{L} \int_{-L}^{L} F_1^k(r_{12}) \, \exp[-\beta \, U(r_{12})] \, \mathrm{d}\xi^x \, \mathrm{d}\xi^y \, \mathrm{d}\xi^z \, ,$$

where r_{12} is the minimum image interatomic distance (2.33). This integral can be written in terms of standard Euclidean distance $r^2 = (\xi^x)^2 + (\xi^y)^2 + (\xi^z)^2$ as

$$f_k = 8\,L^3 \int_0^{L/2}\int_0^{L/2}\int_0^{L/2} F_1^k(r)\,\exp[-\beta\,U(r)]\,\mathrm{d}\boldsymbol{\xi}\,, \tag{2.35}$$

where $\mathrm{d}\boldsymbol{\xi} = \mathrm{d}\xi^x\,\mathrm{d}\xi^y\,\mathrm{d}\xi^z$. In order to analyse f_k further by implementing a far-field approximation, we need to make sure we are in a regime where the integrand is small - we do this by introducing a cut-off γ_c, which will divide the cube $[0, L/2]^3$ into 8 cuboid subdomains, including

$$\begin{aligned} \Omega_1 &= [0,\gamma_c]^3, & \Omega_2 &= [0,\gamma_c]^2 \times [\gamma_c, L/2], \\ \Omega_3 &= [0,\gamma_c] \times [\gamma_c, L/2]^2, & \Omega_4 &= [\gamma_c, L/2]^3. \end{aligned}$$

Utilising the symmetry of the problem, we can rewrite integral (2.35) as

$$f_k = 8\,L^3 \left(\int_{\Omega_1} + \; 3\int_{\Omega_2} + \; 3\int_{\Omega_3} + \int_{\Omega_4} \right) F_1^k(r)\,\exp[-\beta\,U(r)]\,\mathrm{d}\boldsymbol{\xi}\,. \tag{2.36}$$

Considering (2.36) for $k = 0$, the integral over Ω_1 is independent of L and provides a bulk contribution to f_0 that will depend on γ_c. The remaining three terms have integration domains that allow the integrand to be accurately described by a Taylor expansion giving the leading order contribution in the asymptotic limit $L \to \infty$ as $f_0 \propto L^6$, which can be rewritten in terms of the density, n, in the form

$$f_0 \;\propto\; n^{-2} \quad \text{as} \quad n \to 0. \tag{2.37}$$

Considering f_k for $k \neq 0$, the integral over Ω_1 in equation (2.36) is again independent of L. However in the far-field expansion the integrals over Ω_2, Ω_3 and Ω_4 all decay with L due to the force factor. As the integration domain has essentially been transformed

into that of interatomic distances about the three coordinates, when we increase the domain length, the interatomic force necessarily decays to 0. Therefore in the limit $L \to \infty$ the dominant term arises from integrating over Ω_1, and we see that, for $k = 2, 4, 6, 8, \ldots,$

$$f_k \propto n^{-1} \quad \text{as} \quad n \to 0\,. \tag{2.38}$$

This leaves us with the final result that in the low density limit $n \to 0$, combining equation (2.8) with asymptotic expressions (2.37) and (2.38),

$$\alpha_k \propto n^{1-k/2} \quad \text{as} \quad n \to 0\,. \tag{2.39}$$

While this result has been calculated for $N = 2$ interacting atoms, it is also confirmed for larger values of N by estimating the k^{th} standardised moments using MD simulations, as it is shown in Figure 2.5.

2.3.2 MD simulations with N interacting atoms

In this section we present the results from MD simulations of many-body systems in three spatial dimensions using different values of N, including the case $N = 2$ (analyzed in Section 2.3.1). Atoms are subject to pairwise interactions governed by a LJ potential, given in equation (3.25). For each system we use a velocity-Verlet [220] integrator and maintain the system in the canonical ensemble by incorporating a Nosé-Hoover thermostat [167], see Appendix 2.6.

We perform two types of MD simulation studies: those that are used for studying how the number density, n, of a system affects standardised moments, and those that aim to probe temperature dependency. In all cases we utilise a time step $\Delta t = 0.01$. In the case of the simulation with $N = 2$ atoms, we initialise the positions of atoms by setting $\mathbf{q}_1 = \mathbf{0}$ and $\mathbf{q}_2 = (L/2, L/2, L/2)$, whereas for the $N = 8$, 64, 512 atom systems, we choose to initialise these on a uniform cubic lattice. In order to

N	t_{sim}	L_0	n_0
2	10^9	5	0.016
8	10^7	3	1/64
64	10^6	5	1/64
512	10^4	10	1/64

Table 2.1: *The length of MD simulation, t_{sim}, the (smallest) box width, L_0, used for simulations with N atoms and density n_0 for MD simulations with varying temperatures.*

initialise the momentum for the $N = 2, 8, 64, 512$-atom systems, we choose to set $\mathbf{p}_i^{x,y,z} = \sqrt{T}\mathcal{N}(0,1)$, for each $i = 1, ..., 512$, and a new random number is generated for each coordinate and atom. In this way each atom has on average $\langle \mathbf{p}_i^2 \rangle = 3T$ and we satisfy the equipartition theorem. A Box-Muller transform is used to create a normally distributed random number from a uniformly distributed random number we obtain from the FORTRAN 95 in-built psuedorandom number generator.

The MD simulation parameters are summarised in Table 2.1, where t_{sim} is the total simulation time used for calculating the required statistics, which is preceded by the initial simulation of length $t_{\text{sim}}/10$ used for equilibrating the system. When investigating the number density dependence, we perform 20 simulations each with a box width of $L = L_0 \times (6/5)^{i-1}$, where $i = 1, 2, \ldots, 20$ labels the simulation number and L_0 is the smallest cubic box width. We simulate the $N = 8, 64, 512$-atom systems with $L_0 =$ 3, 5, 10, respectively. This enables direct comparison because we can identify triplets of simulated systems corresponding to systems of the same number densities. The two-atom system however is simulated in a sparser regime with $L_0 = 5$.

We calculate statistics on the fly for every time step (after the first 10% initialisation), for every atom and for each coordinate - therefore we average the computed results over the number of time steps ($0.9 \times t_{\text{sim}}/\Delta t$) and atom coordinates ($3N$). In particular, the statistics are calculated over $0.9 \times 3\,N\,t_{\text{sim}}/\Delta t$ data points. This is equal to 6×10^{11} (resp. 1.536×10^9) data points in the simulation with $N = 2$ (resp.

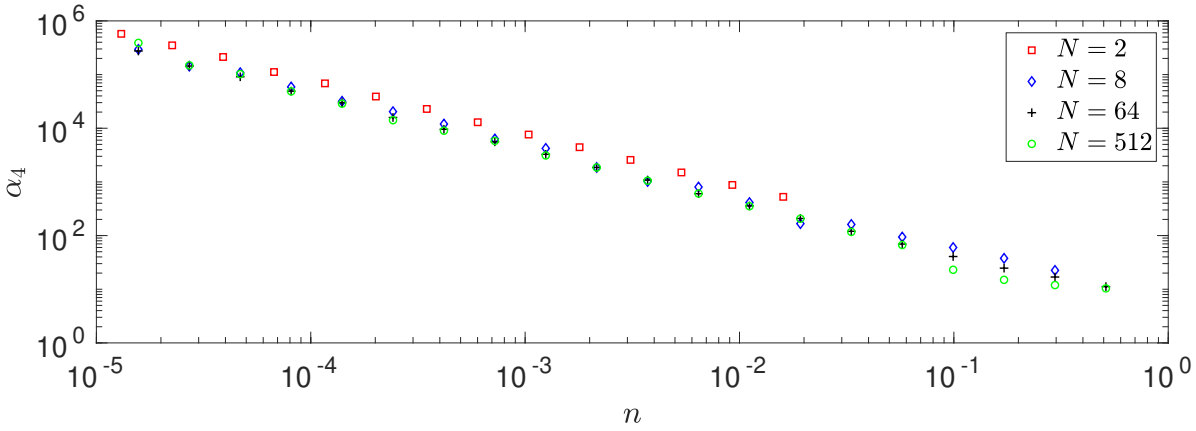


Figure 2.4: *Dependence of kurtosis α_4 on density n. Each of the larger atomic systems ($N = 8, 64, 512$) is simulated over the same domain of number densities, while the $N = 2$ system is simulated in a sparser domain. We truncate the results of the $N = 2$ simulation in the plot, though the additional data points are used in calculating asymptotic results displayed in Figure* 2.5.

$N = 512$) atoms.

Since we use PBCs, we are effectively looking at an infinite fluid with a prescribed number density $n = N/L^3$ and we can calculate the behaviour of kurtosis α_4 as n varies. The results are presented in Figure 2.4. We see general agreement between behaviour of each of the four systems. We see when n is equal, the values of kurtosis are larger for $N = 2$ than for the many-body systems with $N = 8, 64, 512$, which agree well amongst themselves.

The results in Figure 2.4 enable us to test the asymptotic expression (2.39) for $k = 4$ derived in the limit $n \to 0$. Utilising similar log-log plots for MD data, we estimate the power law behaviour of each standardised moment, α_k, for $k = 4, 6, 8, 10, 12$.

Figure 2.5 illustrates the results. All systems agree well with the predicted asymptotic behaviour (2.39), in particular the $N = 512$ atom system. There is a slight deviation between the results due to the fact that the smaller atom systems require a larger t_{sim} in order to converge fully to the predicted value. This discrepancy is amplified when looking at higher standardised moments due to the fact that we are calculating statistics resulting from F_1^{12} (*i.e.* for α_{12}) compared to F_1^4 (*i.e.* for α_4),

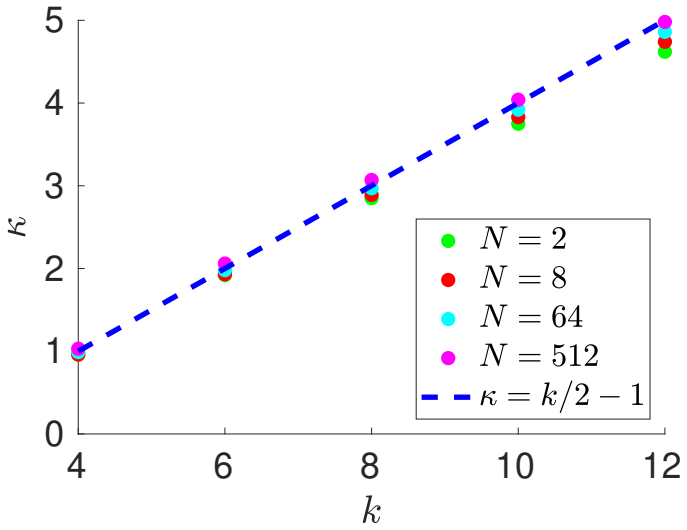


Figure 2.5: *Comparison of the results of N-atom simulations for a range of values of the total number of atoms, N, found within the principal cubic domain. After long time simulation, we compute the asymptotic behaviour $\alpha_k \propto n^{-\kappa}$ and compare the leading order power scalings for each system. We compare this with the theoretical result* (2.39) *(denoted with a blue dashed line) that in the limit $n \to 0$ we expect the universal behaviour $\kappa = k/2 - 1$.*

for example.

The dependence of kurtosis α_4 on temperature T is presented in Figure 2.6, where we keep the density fixed at $n = n_0$ given in Table 2.1. We observe that as temperature increases so does the kurtosis of the force distribution associated with each system. This can be explained in terms of the dynamics of the interacting atom system. If we maintain each system in the canonical ensemble, we expect on average that each atom will have a kinetic energy equivalent to $3\,T/2$ (when in reduced units). As we increase this target temperature, the atoms become more energetic and thus are able to probe closer interatomic distances before a large repulsive force overcomes this inertial attraction. The range of forces on the tagged atom widens as temperature increases and therefore contributes to more outlier results in the distribution - leading to heavier tails and therefore distributions which become increasingly leptokurtic.

In Figure 2.6, we observe that there is a qualitative difference between the results

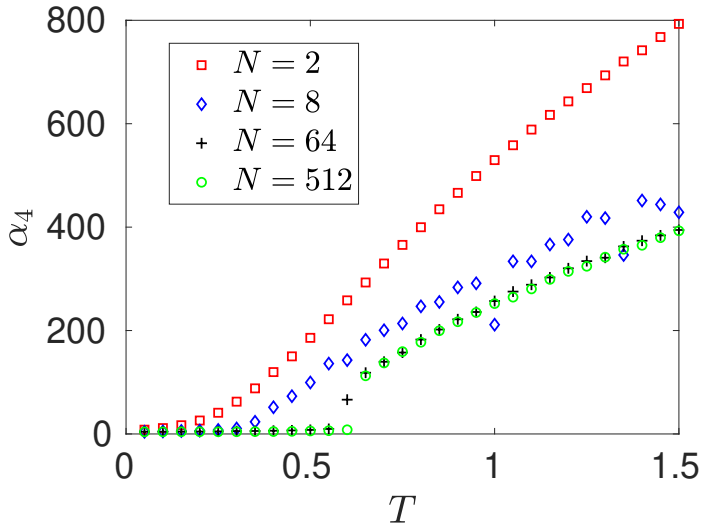


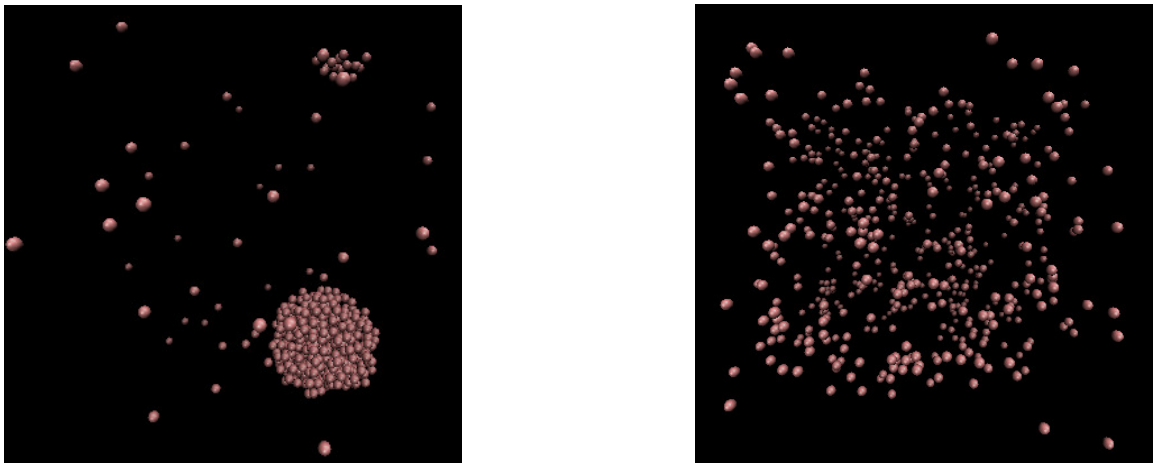
Figure 2.6: *Dependence of kurtosis α_4 on temperature T. Each atomic system is simulated at approximately the same density $n = n_0$ given in Table* 2.1.

for $N = 2$ and larger atom systems. We see a bifurcation for the $N = 64, 512$ system at some critical temperature $T_c \in (0.6, 0.65)$, where a steady increase in kurtosis meets a large jump resulting in a rapid increase. This difference is due to a clustering mechanism which has been seen in MD simulations of LJ fluids [235]. From our results we see that the $N = 2$ system has missed this behaviour completely. Snapshots of the $N = 512$-atom system at some $T = 0.6 < T_c$, and $T = 0.66 > T_c$ are displayed in Figure 2.7. For $T = 0.6$, we see a large cluster has formed in the many-atom system. There would be far fewer outlier force results in this case due to the fact that the large majority of atoms are moving as a collective and effectively have fixed interatomic forces. Compared to the $T = 0.66$ snapshot, where we see that the atoms are too kinetically unstable to form these larger stable cluster structures, this results in more outlier forces felt between atoms due to the fact that the system is intrinsically more disordered. The clustering mechanisms seen (relating to liquid-liquid phase separation) have been studied on dilute LJ fluids [130]; here we see that this clustering results in a bifurcation on standardised moments of the force distribution.

To understand the underlying variations of kurtosis, α_4, with respect to changes in temperature and density, we use 12×16 MD simulations with $N = 512$ atoms, varying simulation parameters (n, T), where $n = 10^{-2} + (i-1)/10$, for $i = 1, 2, \ldots, 12$, and $T = j/10$, for $j = 1, 2, \ldots, 16$. Each simulation was performed with similar initialisation as described previously in Section 2.3.2 with $t_{\text{sim}} = 10^6$. The results for excess kurtosis ($\alpha_4 - 3$) are displayed in Figure 2.8. Here a bifurcation can be seen when using the smallest density $n = 0.01$, as the change in colour is prominent in this vertical strip, indicating a large change of kurtosis. This occurs around $T = 0.6$, which is consistent with the result in Figure 2.6, where we saw the bifurcation similarly located, though the slight shift in temperature is accounted for by the shift in density parameters used in each simulation (namely $n = 0.01$ and $n = 1/64$).

In general, this low density strip contains the largest values of kurtosis, and covers much of the purely gas phase of the LJ fluid. This chapter has so far probed the low density limit in an attempt to understand why the standardised moments of force are so large, though Figure 2.8 gives a good overview that in general, regardless of phase, a decrease in temperature, or an increase in density, systematically lead to a lower value of standardised moments. In this case as $n \to \infty$ or $T \to 0$, we expect the $\alpha_4 \to 3$ (excess kurtosis tends to zero). This limiting regime corresponds to the solid phase of a LJ system, where the force variations are minimal and the distribution is Gaussian. There is not enough space, nor energy, that lead to (many) outlier forces experienced by any atom, so the force distribution increasingly resembles a Gaussian, the deeper we probe in these regions. This intuition was demonstrated analytically in Section 2.2.5 when we showed this limiting behaviour on a 1d cartoon model with equation (2.32).

It is interesting to note that these changes in values of α_4 appear smooth about changes in temperature and density (in absence of the bifurcation point for larger values of n), regardless of phase transitions.



(a) $T < T_c$ (b) $T > T_c$

Figure 2.7: *Snapshots [121] of the MD simulation are taken for the system with* $N = 512$ *atoms at time* $t = 7.5 \times 10^5$ *for:* (a) $T = 0.6 < T_c$*; and* (b) $T = 0.66 > T_c$*. The number density is* $n = 1/64$*.*

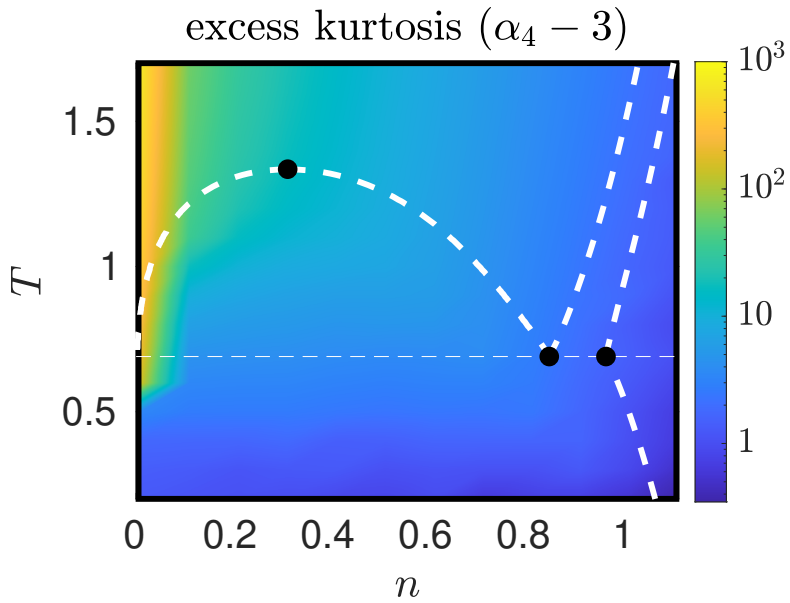


Figure 2.8: *The excess kurtosis,* $\alpha_4 - 3$*, calculated as a function of density* n *and temperature* T *for* $n \leq 1.11$ *and* $T \leq 1.6$*. The white dotted lines describe coexistence lines of different phases of a LJ fluid taken from the literature [193, 206, 205, 157]. The solid black dots indicate (from left to right), the critical point and vapour-liquid-solid triple points.*

2.4 Discussion

In Section 2.2 we have demonstrated use of a variety of methods to study the standardised moments of the force distribution in order to probe both their temperature and number density dependence. This gave way to a rich structure where we show that the partition function for a 1d system can be calculated entirely from these standardised moments. Extending the far-field method introduced in Section 2.2.2 to a system with N atoms in three-dimensional physical space, Section 2.3 studies the dependence of α_k on number density n, deriving the asymptotic expression (2.39). Though results are derived with a two-body system in a principal cubic domain, utilising the minimum image convention implies that we are studying an infinite fluid at a prescribed number density. Therefore analytic results are contrasted with MD simulations of four systems of $N = 2, 8, 64, 512$ interacting LJ atoms and these are compared.

The results agree well with theoretical predictions though the results for systems with larger values of N are seen to converge more readily to the theoretically predicted results. In particular, rich dynamics such as clustering of LJ fluids is completely missed by the systems with smaller values of N, but captured for systems with N as small as $N = 64$ atoms. In general, as temperature increases α_k increases due to the higher kinetic energy of atoms allowing them to push closer together and experience larger forces.

Clustering exhibited at the vapour-liquid coexistence phase incurs a bifurcation point whereby a large increase is seen in the standardised moments of force in Figure 2.6, though a general increase in temperature, or decrease in number density, results in an increase in a_4 regardless of the temperature/number density domain studied, as shown in Figure 2.8.

2.5 Derivation of equations (2.28) and (2.29)

The constants used in equation (2.30), namely $C_{k,1}$ and $C_{k,2}$, are given in equations (2.28) and (2.29), which can be derived in the following manner. Using $F^k(r) = (U'(r))^k$ for even values of k and $U'(r^*) = 0$, we first note that

$$\begin{aligned} F^{k,\,k}(r^*) &= k!\,(U''(r^*))^k\,, \\ F^{k,\,m}(r^*) &= 0, \qquad\qquad \text{for} \quad m \leq k-1, \end{aligned}$$

where $F^{k,\,m}$ denotes the m^{th} derivative of F^k, i.e. the m^{th} derivative of the k^{th} power of F. Therefore, the first three non-zero terms of the Taylor expansion of $F^k(r)$ around $r = r^*$ are

$$\begin{aligned} F^k(r) \;\approx\; & (r-r^*)^k\,(U''(r^*))^k + (r-r^*)^{k+1}\,\frac{F^{k,\,(k+1)}(r^*)}{(k+1)!} \\ & +(r-r^*)^{k+2}\,\frac{F^{k,\,(k+2)}(r^*)}{(k+2)!}\,. \end{aligned} \tag{2.40}$$

Therefore, we have $C_{k,1} = F^{k,\,(k+1)}(r^*)/(k+1)!$ and $C_{k,2} = F^{k,\,(k+2)}(r^*)/(k+2)!$ and, to derive equations (2.28) and (2.29), we need to express derivatives $F^{k,\,m}(r^*)$ for $m = k+1$ and $m = k+2$ in terms of derivatives of $U(r)$ at $r = r^*$. Using the product rule, the m^{th} derivative of $F^k(r)$ can be, in general, written as a finite sum of the form

$$F^{k,\,m}(r) = \sum_{\alpha_0,\alpha_1,\ldots,\alpha_m=0}^{k} C(\alpha_0,\alpha_1,\ldots,\alpha_m)\,\prod_{i=0}^{m}\left(F^{(i)}(r)\right)^{\alpha_i}, \tag{2.41}$$

where $F^{(i)}(r)$ is the i^{th} derivative of function $F(r)$ and $C(\alpha_0,\alpha_1,\ldots,\alpha_m)$ are constants, many of them equal to zero. In fact, all terms in the expansion (2.41) have multiplicities that sum to k, that is we can only sum over sequences satisfying

$$\sum_{i=0}^{m}\alpha_i = k, \tag{2.42}$$

and all terms in the expansion (2.41) have m derivatives, that is, we have

$$\sum_{i=0}^{m} i\,\alpha_i = m, \tag{2.43}$$

where $\alpha_i \in \{0, 1, \ldots, k\}$ for $i = 0, 1, 2, \ldots, m$. Equation (2.43) is of the form of a finite Diophantine equation, which has no closed form for the number of solutions. In particular, simplifying equation (2.41) by solving equations (2.42)–(2.43) is, in general, not possible. However, noting the specific property that $F(r^*) = 0 = U'(r^*)$, we see that all terms that have $\alpha_0 \neq 0$ will vanish when evaluated at this unique minimum $r = r^*$. In particular, we will obtain relatively simple forms of the sum (2.41) for $m = k + 1$ and $m = k + 2$ by considering equations (2.42)–(2.43) with $\alpha_0 = 0$.

First, let us consider that $m = k + 1$. Using $\alpha_0 = 0$, there is only one solution of equations (2.42)–(2.43) in non-negative integers, namely $\alpha_1 = k - 1$, $\alpha_2 = 1$ and $\alpha_3 = \alpha_4 = \cdots = 0$. Therefore, equation (2.41) implies

$$F^{k,\,(k+1)}(r^*) = C(0, k-1, 1, 0, \ldots, 0)\,\left(F^{(1)}(r^*)\right)^{k-1}\,F^{(2)}(r^*).$$

Using the general Leibniz rule [213], we evaluate the combinatorial prefactor as $C(0, k-1, 1, 0, \ldots, 0) = k(k+1)!/2$. Substituting into $C_{k,1} = F^{k,\,(k+1)}(r^*)/(k+1)!$ and using $F(r^*) = -U'(r^*)$ and that k is an even integer, we obtain formula (2.28).

Second, we will consider the case $m = k+2$. Using $\alpha_0 = 0$, there are two solutions of equations (2.42)–(2.43) in non-negative integers. The first solution is $\alpha_1 = k - 1$, $\alpha_2 = 0$, $\alpha_3 = 1$ and $\alpha_4 = \alpha_5 = \cdots = 0$. The second solution is $\alpha_1 = k-2$, $\alpha_2 = 2$ and $\alpha_3 = \alpha_4 = \cdots = 0$. Therefore, equation (2.41) implies

$$\begin{aligned} &F^{k,\,(k+2)}(r^*) \\ &\quad= C(0, k-1, 0, 1, 0, \ldots, 0)\,\left(F^{(1)}(r^*)\right)^{k-1}\,F^{(3)}(r^*) \\ &\quad+ C(0, k-2, 2, 0, 0, \ldots, 0)\,\left(F^{(1)}(r^*)\right)^{k-2}\left(F^{(2)}(r^*)\right)^{2}. \end{aligned}$$

Using the general Leibniz rule [213], we evaluate these combinatorial prefactors as

$$\begin{aligned} C(0,k-1,0,1,0,\ldots,0) &= \frac{k}{6}\,(k+2)!, \\ C(0,k-2,2,0,0,\ldots,0) &= \frac{k(k-1)}{8}\,(k+2)!. \end{aligned}$$

Substituting into formula $C_{k,2} = F^{k,\,(k+2)}(r^*)/(k+2)!$ and using $F(r^*) = -U'(r^*)$ and that k is an even integer, we obtain equation (2.29). Thus, we have arrived at the the expressions for $C_{k,2}$ and $C_{k,2}$ that are used in equation (2.30).

2.6 Thermostats used in MD simulations

Considering 3d simulations in Section 2.3.2, we use a Nosé-Hoover thermostat. Its parameter, originally [167] denoted Q, is the relaxation time of the thermostat. It is a measure of how strongly the thermostat is attached to the dynamics of the system. We choose a cautious value of $Q = 10\,T$ for each simulation; this linear scaling with T is necessary as we need to more tightly couple the thermostat at lower temperatures in order to accurately maintain the system in the canonical ensemble [122]. This linear scaling in temperature can be understood from the following equation for auxiliary variable ξ:

$$\frac{d\xi}{dt} = Q^{-1}(\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} - N_d k_B T).$$

Here one would naturally like to pick Q to have a linear scaling of temperature (or energy) from dimensional analysis.

For 1d simulations in Section 2.2, we maintain the canonical ensemble at a target (reduced) temperature T by implementing a Langevin thermostat. This is due to problems with ergodicity utilising the Nosé-Hoover thermostat for small systems [215, 216]. Here the evolution of the free atom is modelled (in reduced units) as [191, 143]

$$\ddot{x} = -\frac{\mathrm{d}U}{\mathrm{d}x} - \tilde{\gamma}\,\dot{x} + \sqrt{2\,\tilde{\gamma}\,T}\,R(t)\,, \tag{2.44}$$

where $R(t)$ is standard white noise, and $\tilde{\gamma}$ acts as a friction parameter. We choose $\tilde{\gamma} = 0.1$ when calculating our illustrative results presented in Section 2.2.

Chapter 3

On the equivalence between the configuration space and force space in simple liquids

3.1 Introduction

In this chapter, we present work on force inversion introduced in Section 1.4. The work is a modified version of a paper being produced in collaboration with Radek Erban. This chapter differs from the paper in that content from the introduction of the paper has been merged with the introduction of the thesis while some additional material, in Sections 3.2.3 and 3.5, are included to incorporate general harmonic interactions and 2-body inversion respectively.

We start with the inverse mapping defined in Section 1.4 given by equation (1.9). Given the instantaneous forces $\mathbf{F} \in \mathbb{R}^{3N}$, and the known pair potential $U : [0, \infty) \to \mathbb{R}$, the inverse mapping (1.9) aims to find a configuration $\mathbf{q} \in \mathcal{X}$ of N atoms that satisfies $\mathbf{F} = F(\mathbf{q})$. To do this, we answer the following three questions:

(Q1) Is $\mathbf{F} \in \mathbb{R}^{3N}$ in the force space $\mathcal{F}$?

(Q2) How many different solutions (up to translation) does the equation $\mathbf{F} = F(\mathbf{q})$ have?

(Q3) How do we calculate a point $\mathbf{q} \in \mathcal{X}$ satisfying $\mathbf{F} = F(\mathbf{q})$?

The answers to questions (Q1)–(Q3) depend on the given pair potential $U : [0, \infty) \to \mathbb{R}$ and on the configuration space $\mathcal{X}$. In Section 3.2.1, we will answer these questions in the case of harmonic potential and $\mathcal{X} = \mathbb{R}^{3N}$. In this case, we have

$$\mathcal{F} = \left\{ \mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_N) \in \mathbb{R}^{3N} \;\middle|\; \sum_{i=1}^{N} \mathbf{F}_i = 0 \right\}, \tag{3.1}$$

that is, one can answer question (Q1) for a given $\mathbf{F} \in \mathbb{R}^{3N}$ by checking that the sum of all forces on individual atoms is zero. In general, the configuration $\mathbf{q}$ which corresponds to given $\mathbf{F} \in \mathcal{F}$ is not unique, because any translation applied to $\mathbf{q}$ will give another solution. In Section 3.2.1, we will answer question (Q2) by showing that $\mathbf{F} = F(\mathbf{q})$ has (up to translation) a unique solution for any $\mathbf{F} \in \mathcal{F}$, which can be found by solving a set of linear equations by Gaussian elimination to answer question (Q3) in the case of harmonic potential and $\mathcal{X} = \mathbb{R}^{3N}$.

Another important case, $\mathcal{X} = \Omega^N$, where Ω is a cuboid equipped with periodic boundary conditions (PBCs), is studied in Section 3.2.2 for the harmonic potential. Here $\mathcal{F}$ is a bounded convex subset of the set given in equation (3.1). We can answer question (Q1) by checking that $\mathbf{F} \in \mathbb{R}^N$ satisfies a set of inequalities (see equation (3.8)). However, unlike in the case of $\mathcal{X} = \mathbb{R}^{3N}$ studied in Section 3.2.1, the solution to the equation $\mathbf{F} = F(\mathbf{q})$ is no longer unique (up to translation), although calculation of a point $\mathbf{q} \in \mathcal{X}$ satisfying $\mathbf{F} = F(\mathbf{q})$ in question (Q3) can still be found in relatively straightforward way using Gaussian eliminations as discussed in Section 3.2.2 for finite domains with PBCs.

Considering general pair potential $U : [0, \infty) \to \mathbb{R}$, the mapping $F(\mathbf{q})$ is nonlinear and the problem of finding its inverse (1.9) leads to finding roots of the system of

nonlinear equations. Then question (Q3) can, in principle, be answered by applying iterative numerical methods, like Newton's method. However, such approaches suffer from extreme sensitivity to the choice of initial condition [93]. In order to overcome this common deficiency in iterative methods, continuation methods have been developed as mentioned in Section 1.5. In Section 3.3, we present the Newton homotopy continuation method (NHCM) [232, 1, 142, 95] that uses analytically obtained solution from Section 3.2 as a starting point for answering questions (Q1)–(Q3) for general potentials and general configuration spaces $\mathcal{X}$. Our illustrative simulations are conducted using the standard Lennard-Jones (LJ) potential [128] in Section 3.4.

To get further insight into this force inversion, the case $N = 2$ is investigated in Section 3.5, where we can solve for the configurations directly in each case. In the higher dimensions ($N > 2$), only the harmonic case can be analytically solved and the general pair potential $U : [0, \infty) \to \mathbb{R}$ case relies on the NHCM calculations. In Section 3.4.1, we discuss how convergence of the algorithms is affected by parameters which: control how quickly one maps from the initial system ($\gamma = 0$) to the final system ($\gamma = 1$), give the number of steps in Newton's method for each fixed γ, and dampen the harmonic constants. Section 3.5 details the $N = 2$ atom inversion, highlighting how to deal with multiple configuration choices. The work is concluded in the final Section 3.6 where results are summarised.

3.2 Calculation of mapping (1.9) for harmonic oscillators

Consider a system of N atoms which interact according to harmonic interactions in domain $\Omega \subset \mathbb{R}^3$, where two important cases are studied: (i) $\Omega = \mathbb{R}^3$ in Section 3.2.1; and (ii) Ω is a cuboid with PBCs in Section 3.2.2. In both cases, the positions of atoms are denoted as $\mathbf{q}_i \in \Omega$, for $i = 1, 2, \ldots, N$ and the pair potential is given by

$$U(r_{ij}) = \frac{k}{2} r_{ij}^2, \tag{3.2}$$

where k is a constant and r_{ij} is the distance between the i^{th} and j^{th} atoms.

3.2.1 The case $\mathcal{X} = \mathbb{R}^{3N}$

If $\Omega = \mathbb{R}^3$, then we have $\mathcal{X} = \mathbb{R}^{3N}$ and the displacement vector between the i^{th} and j^{th} atoms is $\boldsymbol{\Delta}_{ji} = \mathbf{q}_i - \mathbf{q}_j$. Substituting potential (3.2) into equation (1.8), we get

$$\mathbf{F}_i = -\sum_{j=1, j\neq i}^{N} k\,\boldsymbol{\Delta}_{ji}. \tag{3.3}$$

Consequently, we can write the mapping $F : \mathcal{X} \to \mathcal{F}$ as the linear mapping $F(\mathbf{q}) = \underline{\underline{A}}\, k\, \mathbf{q}$, where $\underline{\underline{A}}$ is a $3N \times 3N$ matrix given by

$$\underline{\underline{A}} = \begin{pmatrix}
(1-N) & 0 & 0 & 1 & 0 & 0 & \dots & 1 & 0 & 0 \\
0 & (1-N) & 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 \\
0 & 0 & (1-N) & 0 & 0 & 1 & \dots & 0 & 0 & 1 \\
1 & 0 & 0 & (1-N) & 0 & 0 & \dots & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & (1-N) & 0 & \dots & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & (1-N) & \dots & 0 & 0 & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0 & 0 & 1 & 0 & 0 & \dots & (1-N) & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 & (1-N) & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & \dots & 0 & 0 & (1-N)
\end{pmatrix}.$$

The nullspace of this matrix has dimension 3 with basis

$$\begin{aligned}
\mathbf{v}_1 &= (1,0,0,1,0,0,\dots,1,0,0), \\
\mathbf{v}_2 &= (0,1,0,0,1,0,\dots,0,1,0), \\
\mathbf{v}_3 &= (0,0,1,0,0,1,\dots,0,0,1).
\end{aligned} \tag{3.4}$$

Applying the Fredholm alternative [89] to $\underline{\underline{A}}\, k\, \mathbf{q} = \mathbf{F}$, for this equation to have solutions $\mathbf{q}$, it must be that any vector in the nullspace of matrix $\underline{\underline{A}}$ in (3.4) (*i.e.* $\mathbf{v}_i^T \underline{\underline{A}} = \mathbf{0}$) satisfies $\mathbf{v}_i^T \mathbf{F} = \mathbf{0}$ for $i = 1, 2, 3$. In which case we obtain (3.1). Moreover, the Fredholm alternative also implies that, for a given $\mathbf{F} \in \mathcal{F}$, any two solutions of $\mathbf{F} = F(\mathbf{q})$ have to differ by a vector in the nullspace of A, given as $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$ where c_1, c_2 and c_3 are constants. That is, they differ up to a translation in the direction of vector (c_1, c_2, c_3). To calculate a solution for given $\mathbf{F} \in \mathcal{F}$, we can choose $\mathbf{q}_1 = (0, 0, 0)$ and calculate $\mathbf{q}_2$, $\mathbf{q}_3$, $\ldots$, $\mathbf{q}_N$ by Gaussian elimination.

3.2.2 Finite domain with periodic boundary conditions

Simulations of simple liquids are often done at constant average density by simulating N atoms in a finite domain with PBCs. In this case, domain Ω can be chosen as a cuboid

$$\Omega = [0, L_x] \times [0, L_y] \times [0, L_z], \tag{3.5}$$

where $L_z > 0$, $L_y > 0$ and $L_z > 0$. Then equation (3.3) is still applicable if we use the PBCs (minimum image convention) to redefine the displacement vector between the i^{th} and j^{th} atoms as below:

$$\boldsymbol{\Delta}_{ji} = \mathbf{q}_i - \mathbf{q}_j - B \left\lfloor B^{-1} \left(\mathbf{q}_i - \mathbf{q}_j\right) \right\rceil, \qquad \text{where} \qquad B = \begin{pmatrix} L_x & 0 & 0 \\ 0 & L_y & 0 \\ 0 & 0 & L_z \end{pmatrix} \tag{3.6}$$

and the function $\lfloor \cdot \rceil$ rounds a real number to the nearest integer (*i.e.* $\lfloor \xi \rceil$ is the unique integer lying in interval $(\xi - 1/2, \xi + 1/2]$ for any real number ξ). In the case of the finite domain (3.5), the force space $\mathcal{F}$ is no longer given by equation (3.1), because all forces in $\mathcal{F}$ are bounded. In particular, if we define the scaled L^∞-norm of a vector

$\mathbf{v} = (v_x, v_y, v_z) \in \mathbb{R}^3$ by

$$\| \mathbf{v} \|'_\infty = \max \left\{ \frac{v_x}{L_x}, \frac{v_y}{L_y}, \frac{v_z}{L_z} \right\}, \tag{3.7}$$

then $\mathcal{F}$ can be written as

$$\mathcal{F} = \left\{ \mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_N) \in \mathbb{R}^{3N} \;\middle|\; \sum_{i=1}^{N} \mathbf{F}_i = 0 \;\text{ and }\; \| \mathbf{F}_i - \mathbf{F}_j \|'_\infty \leq \frac{Nk}{2} \text{ for all } i, j \right\}, \tag{3.8}$$

which generalizes equation (3.1) to the finite domain (3.5) with PBCs. We observe that $\mathcal{F}$ in equation (3.8) is a proper subset of the set (3.1) obtained in Section 3.2.1. Therefore, given $\mathbf{F} \in \mathcal{F}$, we can use the approach described in Section 3.2.1 (based on Gaussian elimination) to obtain $\mathbf{q} \in \mathbb{R}^{3N}$ satisfying $\mathbf{F} = F(\mathbf{q})$ for the case of $\Omega = \mathbb{R}^3$. However, the additional conditions on $\mathbf{F} \in \mathcal{F}$, namely the inequalities $\| \mathbf{F}_i - \mathbf{F}_j \|'_\infty \leq Nk/2$ for all $i \neq j$, imply that the obtained $\mathbf{q}$ satisfies

$$\max_{i \in \{1,2,\ldots,N\}} q_{i,x} \;\leq\; \frac{L_x}{2} + \min_{i \in \{1,2,\ldots,N\}} q_{i,x}, \qquad \max_{i \in \{1,2,\ldots,N\}} q_{i,y} \leq \frac{L_y}{2} + \min_{i \in \{1,2,\ldots,N\}} q_{i,y},$$
$$\text{and} \qquad \max_{i \in \{1,2,\ldots,N\}} q_{i,z} \leq \frac{L_z}{2} + \min_{i \in \{1,2,\ldots,N\}} q_{i,z}.$$

In particular, we can apply a translation to get the solution $\overline{\mathbf{q}}$ of equation $\mathbf{F} = F(\mathbf{q})$ satisfying

$$\overline{\mathbf{q}}_i \in \left[-\frac{L_x}{4}, \frac{L_x}{4} \right] \times \left[-\frac{L_y}{4}, \frac{L_y}{4} \right] \times \left[-\frac{L_z}{4}, \frac{L_z}{4} \right], \qquad \text{for all} \quad i = 1, 2, \ldots, N. \tag{3.9}$$

Then, question (Q3) can be answered using the same algorithm as in Section 3.2.1.

While the obtained solution has been unique (up to translation) in Section 3.2.1, this is no longer the case for the finite cuboid domain (3.5) with PBCs and the answer to question (Q2) will depend on the value of N. For example, suppose that the number of atoms, N, is the product of three odd integers N_x, N_y and N_z, that is,

$N = N_x N_y N_z$. Then we can arrange atoms on a cubic $N_x \times N_y \times N_z$ lattice, which results in configuration $\mathbf{q}^0$ given by

$$\mathbf{q}^0_{(i-1)N_x^2+(j-1)N_y+\ell} = \left[\frac{(i-1/2)\,L_x}{N_x}, \frac{(j-1/2)\,L_y}{N_y}, \frac{(\ell-1/2)\,L_z}{N_z}\right], \tag{3.10}$$

for $i = 1, 2, \ldots, N_x$, $j = 1, 2, \ldots, N_y$, and $\ell = 1, 2, \ldots, N_z$. The configuration $\mathbf{q}^0$ corresponds to the zero force vector $\mathbf{F} = (0, 0, \ldots, 0)$, but the zero force vector can also be obtained if we (for any odd N) arrange atoms at regular intervals on a line parallel to any line connecting two vertices of the cuboid Ω, or if we arrange atoms on a square two-dimensional lattice in a plane parallel to one of the sides of Ω. However, none of these solutions is equal to the solution (3.9) for $\mathbf{F} = (0, 0, \ldots, 0)$, which is (up to translation) equal to $\overline{\mathbf{q}} = \mathbf{F} = (0, 0, \ldots, 0)$.

The solution $\overline{\mathbf{q}}$ is, in general, located in a small subdomain of Ω, which is illustrated in Figure 3.1, where we present an example with $N = 11^3 = 1331$ atoms. Their positions have been randomly sampled using the uniform distribution in cuboid domain Ω and we plot the projections of their positions to the x-y plane in Figure 3.1(a) as blue circles. The corresponding force vector $\mathbf{F}$ is plotted in Figure 3.1(b). Next, if we solve the equation $\mathbf{F} = F(\mathbf{q})$ by Gaussian elimination with the force vector $\mathbf{F}$ plotted in Figure 3.1(b), we do not obtain the configuration we started with, because the obtained solution (after translation) satisfies the condition (3.9), which restricts the positions of atoms into a subdomain containing only $1/8^{\text{th}}$ of the volume of the domain Ω. In fact, the calculated positions of the atoms in the configuration $\overline{\mathbf{q}}$ (translated to the middle of the domain Ω) are plotted as the red dots in Figure 3.1(a) and they can all be fit into a small cubic subdomain of Ω, which has less than 0.004% of the volume of Ω. Both red dots and blue circles correspond to the same force vector $\mathbf{F}$, but the calculated configuration $\overline{\mathbf{q}}$ leaves relatively large vacuum spaces in domain

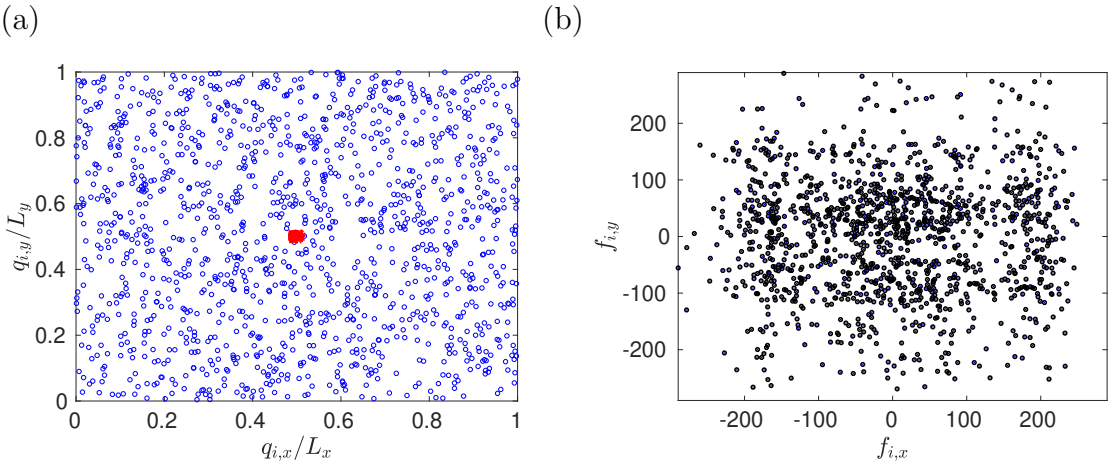


Figure 3.1: (a) *The positions* $\mathbf{q}_i$, $i = 1, 2, \ldots, N$, *of* $N = 1331 = 11^3$ *atoms which have been sampled from the uniform random distribution in the cuboid domain* Ω *given by* (3.5) *(blue circles). We plot their scaled coordinates* $q_{i,x}/L_x$ *and* $q_{i,y}/L_y$ *in the x-y plane. The solution* $\overline{\mathbf{q}}$ *of equation* $\mathbf{F} = F(\overline{\mathbf{q}})$ *given by Gaussian elimination in equation* (3.9)*, where* $\mathbf{f}$ *is given in panel* (b)*, are plotted (after translation to the middle of the domain* Ω*) as red dots.*
(b) *The forces* $\mathbf{F}_i$, $i = 1, 2, \ldots, 121$, *corresponding to the configuration in panel* (a)*. We plot the x-coordinate of each force,* $f_{i,x}$*, against the corresponding y-coordinate,* $f_{i,y}$*.*

Ω. To avoid this in what follows, we use the configuration $\mathbf{q}$ obtained by

$$\mathbf{q} = \overline{\mathbf{q}} + \mathbf{q}^0, \tag{3.11}$$

where $\overline{\mathbf{q}}$ and $\mathbf{q}^0$ are given by (3.9) and (3.10), respectively. In particular, we assume that N can be written as the product of three odd numbers, $N = N_x N_y N_z$, as we have done in Figure 3.1, where $N_x = N_y = N_z = 11$. Then $\mathbf{q}^0$ is well-defined by equation (3.10) and satisfies $\mathbf{0} = F(\mathbf{q}^0)$, while $\overline{\mathbf{q}}$ defined by equation (3.9) satisfies $\mathbf{F} = F(\overline{\mathbf{q}})$. Moreover, we will assume that $\overline{\mathbf{q}}$ is sufficiently close to $\mathbf{0}$ so that $\mathbf{q}$ given by equation (3.11) is a solution to equation $\mathbf{F} = F(\mathbf{q})$. A sufficient condition for this is to assume that $\mathbf{F} \in \mathcal{F}$ satisfies

$$\| \mathbf{F}_i - \mathbf{F}_j \|'_\infty \leq \frac{Nk}{2 \max\{N_x, N_y, N_z\}}, \qquad \text{for all } i = 1, 2, \ldots, N, j = 1, 2, \ldots, N, \tag{3.12}$$

where $\|\cdot\|'_\infty$ is the scaled L^∞-norm given by definition (3.7).

3.2.3 Generalized harmonic interactions

While we are interested in studying simple liquids where all interactions are governed by the same pair potential $U : [0, \infty) \to \mathbb{R}$, our conclusions in Section 3.2.1 will be valid even if use different values of spring constants between different atoms. That is, we consider pairwise interactions governed by the harmonic potential with constants k_{ij}

$$U(r_{ij}) = \frac{k_{ij}}{2} r_{ij}^2. \tag{3.13}$$

Then the equations (1.8) and (3.3) generalize to

$$\mathbf{F}_i = -\sum_{j=1, j\neq i}^{N} k_{ij}\, \Delta_{ji}. \tag{3.14}$$

The mapping $F : \mathcal{X} \to \mathcal{F}$ can again be written as a linear mapping in matrix form, where the nullspace of the matrix is spanned by vectors (3.4). In particular, we can again apply the Fredholm alternative to show the force space $\mathcal{F}$ is given by equation (3.1) and the solutions to $\mathbf{F} = F(\mathbf{q})$ are unique up to translations.

We note that in the case with no PBCs, we can use the first stage of Algorithm 1, the harmonic inversion, with equations (3.13) and (3.14), as illustrated in Table 3.1.

3.3 Newton homotopy continuation method

We consider a general (differentiable) pair potential $U : [0, \infty) \to \mathbb{R}$ that defines forces on atoms using equation (1.8). An example is presented in Section 3.4, where the pair potential U is given as the LJ potential (3.25). Our goal is to design a computational approach to answer question (Q3) for the given potential $U : [0, \infty) \to \mathbb{R}$. In particular, in what follows, we assume that our input force vector $\mathbf{F} \in \mathbb{R}^{3N}$ is in the corresponding force space $\mathcal{F}$: $\mathbf{F} \in \mathbb{R}^{3N}$ and has to satisfy $\sum_{i=1}^{N} \mathbf{F}_i = 0$.

Harmonic mixtures for $N = 4$		
$\mathbf{k}^{\mathrm{HM}}$	(n_1, n_2) for Algorithm 1	$\mathbf{q}$
$(0.5, 0.5, 0.75, 0.75, 0.75, 1)^T$	$(12, 17)$	$(0, 0, 0, -0.671, -0.859, -0.158, 0.036, 0.416, -1.016, 0.115, 0.368, 1.028)^T$
$(0.3, 0.3, 0.5, 0.5, 0.5, 0.75)^T$	$(15, 10)$	$(0, 0, 0, -2.299, -0.754, -1.735, -1.573, -0.063, -1.262, -1.052, 0.146, -0.322)^T$
$(1, 1, 1.25, 1.25, 1.25, 1.5)^T$	$(24, 20)$	$(0, 0, 0, -0.682, -0.685, -0.542, -0.177, 0.1614, -1.077, 0.0923, 0.337, 1.042)^T$

Table 3.1: *In this table we provide a set of spring constants k_{ij} for a harmonic mixture of atoms: two of type A and two of type B, and the consistent configurations that arise in $\Omega = \mathbb{R}^{3N}$ after using the NHCM detailed in Algorithm* 1. *For the case $N = 4$, where we denote atom* 1 *and* 2 *as type A,* 3 *and* 4 *as type B:* $\mathbf{k} = (A-A, A-B, A-B, A-B, A-B, B-B)$ *for $X - Y$ harmonic interactions.*

That is, $\mathbf{F} \in \mathbb{R}^{3N}$ is in the force space $\mathcal{F}$ obtained in equation (3.1) for the harmonic potential. Moreover, $\mathbf{F} \in \mathbb{R}^{3N}$ is in the force space $\mathcal{F}$ obtained in equation (3.8) for the harmonic potential with PBCs in the finite domain provided that we choose the spring constant k sufficiently large. In either case, we can use the Gaussian elimination to find a solution to equation $\mathbf{F} = F(\mathbf{q})$ for the harmonic potential, as discussed in Section 3.2.

The NHCM uses a one-parameter family of pair potentials given as

$$U_\gamma(r_{ij}) = (1-\gamma)\, U_0(r_{ij}) + \gamma\, U(r_{ij}), \quad \text{where} \quad \gamma \in [0, 1] \tag{3.15}$$

where U_0 is a potential for which the answers to questions (Q1), (Q2) and (Q3) are known. In this work, we choose U_0 as the harmonic potential (3.2) which is

characterized by one parameter, k_0, i.e.

$$U_0(r_{ij}) = \frac{k_0}{2}\, r_{ij}^2,$$

and for which we have answered questions (Q1), (Q2) and (Q3) in Section 3.2.1 by using the Fredholm alternative and Gaussian elimination.

The parameter $\gamma = [0, 1]$ in equation (3.15) is incrementally increased to solve a system of nonlinear equations derived for each $\gamma > 0$, starting from a fully harmonic potential, U_0, for $\gamma = 0$, and finishing with a configuration fully described by the pair potential $U : [0, \infty) \to \mathbb{R}$. The total force on atom i is given by equation (1.8) as the sum of contributions from pairwise forces from atoms $j \neq i$. It can be rewritten as

$$\mathbf{F}_i = -\nabla \left(\sum_{j=1, j\neq i}^{N} U\big(r_{ij}\big) \right) = - \sum_{j=1, j\neq i}^{N} k(r_{ij})\, \boldsymbol{\Delta}_{ji}, \tag{3.16}$$

where $\boldsymbol{\Delta}_{ji} = \mathbf{q}_i - \mathbf{q}_j$ for $\Omega = \mathbb{R}^{3N}$ or $\boldsymbol{\Delta}_{ji}$ is defined by equation (3.6) for cuboid domain (3.5) with PBCs, and function $k : [0, \infty) \to \mathbb{R}$ depends on the given pair potential $U : [0, \infty) \to \mathbb{R}$ as

$$k(r) = \frac{1}{r} \frac{\mathrm{d}U}{\mathrm{d}r}(r). \tag{3.17}$$

Considering $\gamma \in [0, 1]$, the answer to question (Q3) for our auxiliary potential U_γ, defined in equation (3.15), can be rewritten as the solution to the following system of nonlinear equations

$$\mathbf{H}_i(\boldsymbol{\Delta}_\gamma) = \mathbf{F}_i + \sum_{j=1, j\neq i}^{N} \big((1-\gamma)\, k_0 + \gamma\, k(r_{ij}) \big) \boldsymbol{\Delta}_{ji} = \mathbf{0} \in \mathbb{R}^3 \quad \text{for} \quad i = 1, 2, \ldots, N-1, \tag{3.18}$$

where our unknown vector $\boldsymbol{\Delta}_\gamma \in \mathbb{R}^{3N-3}$ is defined by

$$\boldsymbol{\Delta}_\gamma = \big(\boldsymbol{\Delta}_{21}, \boldsymbol{\Delta}_{32}, \boldsymbol{\Delta}_{43}, \ldots, \boldsymbol{\Delta}_{N-1\ N-2}, \boldsymbol{\Delta}_{N\ N-1} \big). \tag{3.19}$$

Algorithm 1 *The NHCM to solve* $\mathbf{F} = F(\mathbf{q})$ *for given force data* $\mathbf{F} \in \mathcal{F}$

Solve $\mathbf{F} = F(\mathbf{q})$ for $\gamma = 0$ by using the Gaussian elimination. The resulting configuration is denoted as $\boldsymbol{\Delta}^{0,n_2+1} \in \mathbb{R}^{3N-3}$.
Set $\Delta\gamma = 1/n_1$.
for $i = 1, 2, \ldots, n_1$ **do**
  Set $\gamma = i\,\Delta\gamma$ and $\boldsymbol{\Delta}^{i,1} = \boldsymbol{\Delta}^{i-1,n_2+1} \in \mathbb{R}^{3N-3}$.
  for $j = 1, 2, \ldots, n_2$ **do**
    Use equation (3.22) to calculate configuration $\boldsymbol{\Delta}^{i,j+1}$.
  end for
end for

Note that we do not define equation (3.18) for $i = N$ because Newton's third law implies $\mathbf{F}_N = -\sum_{j=1}^{N-1} \mathbf{F}_j$. That is, we are calculating a configuration $\mathbf{q} \in \mathcal{X}$ up to translation and we only have $3N - 3$ degrees of freedom. If we consider the case $\Omega = \mathbb{R}^3$, the displacements $\boldsymbol{\Delta}_{ji}$ for $j - i \neq 1$ can be written in terms of our unknown variables (3.19) as

$$\boldsymbol{\Delta}_{ji} = \boldsymbol{\Delta}_{j\ j-1} + \boldsymbol{\Delta}_{j-1\ j-2} + \cdots + \boldsymbol{\Delta}_{i+1\ i}. \tag{3.20}$$

Substituting into equation (3.18), we obtain

$$\begin{aligned} \mathbf{H}_i(\boldsymbol{\Delta}_\gamma) = \mathbf{F}_i \ &- \ \sum_{\ell=2}^{i} \boldsymbol{\Delta}_{\ell\,\ell-1} \left[(\ell-1)(1-\gamma)\, k_0 + \gamma \sum_{j=1}^{\ell-1} k(r_{i\,j}) \right] \\ &+ \ \sum_{\ell=i+1}^{N} \boldsymbol{\Delta}_{\ell\,\ell-1} \left[(N-\ell+1)(1-\gamma)\, k_0 + \gamma \sum_{j=\ell}^{N} k(r_{j\,i}) \right] = \mathbf{0} \in \mathbb{R}^3. \end{aligned} \tag{3.21}$$

Defining $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_{N-1}) \in \mathbb{R}^{3N-3}$, we solve the nonlinear system $\mathbf{H}(\boldsymbol{\Delta}_\gamma) = \mathbf{0} \in \mathbb{R}^{3N-3}$ for unknown $\boldsymbol{\Delta}_\gamma \in \mathbb{R}^{3N-3}$ using the NHCM specified as Algorithm 1, where we increase γ linearly in steps of $\Delta\gamma = 1/n_1$. Here, n_1 is one of the parameters of our NHCM approach. For each value of $\gamma = i\,\Delta\gamma$, where $i = 1, 2 \ldots, n_1$, we use n_2 iterations of Newton's method to calculate $\boldsymbol{\Delta}_\gamma$. Namely, we calculate

$$\boldsymbol{\Delta}^{i,j+1} = \boldsymbol{\Delta}^{i,j} - J^{-1}(\boldsymbol{\Delta}^{i,j})\,\mathbf{H}, \qquad \text{for } j = 1, 2, \ldots, n_2, \tag{3.22}$$

where $J : \mathbb{R}^{(3N-3)\times(3N-3)}$ is the Jacobian of $\mathbf{H}$. At the end of each Newton's method, the last configuration is stored and used as an initial guess for the subsequent Newton's method by specifying the initial guess in equation (3.22) as $\mathbf{\Delta}^{i,1} = \mathbf{\Delta}^{i-1,n_2+1} \in \mathbb{R}^{3N-3}$. To calculate the Jacobian in equation (3.22), we note that

$$\begin{aligned}
\frac{\partial \mathbf{H}_i}{\partial \mathbf{\Delta}_{p\ p-1}} &= -\left[\delta_{\ell p}\,(\ell-1)\,(1-\gamma)k_0\,\mathbb{I}_{3x3} + \delta_{\ell p}\sum_{j=1}^{\ell-1}\gamma\,k(r_{i\,j})\,\mathbb{I}_{3x3}\right]\Theta(i+1-p) \\
&\quad + \left[\delta_{\ell p}\,(N-\ell+1)(1-\gamma)k_0\,\mathbb{I}_{3x3} + \delta_{\ell p}\sum_{j=\ell}^{N}\gamma\,k(r_{j\,i})\,\mathbb{I}_{3x3}\right]\Theta(p-i) \\
&\quad - \sum_{\ell=2}^{i}\left[\Delta_{\ell\,\ell-1}\left[\sum_{j=1}^{\ell-1}\gamma a_{i\,j}(\delta_{p\,j+1}+\delta_{p\,j+2}+\cdots+\delta_{p\,i})\Delta_{ij}\right]\right] \\
&\quad + \sum_{\ell=i+1}^{N}\left[\Delta_{\ell\,\ell-1}\left[\sum_{j=\ell}^{N}\gamma a_{j\,i}(\delta_{p\,i+1}+\delta_{p\,i+2}+\cdots+\delta_{p\,j})\Delta_{ji}\right]\right],
\end{aligned} \tag{3.23}$$

where $i = 1, \ldots, N-1$ and $p = 2, 3, \ldots, N$, δ_{lp} is the Kronecker delta function and $\mathbb{I}_{3x3}$ is the 3x3 identity matrix.

The factors of $a_{i\,j}$ that appear are defined by (where $i > j$):

$$\frac{\partial k(r_{ij})}{\partial\left(\Delta_{p\,p-1}\right)_{x,y,z}} = a_{i\,j}\left(\delta_{p\,j+1}+\delta_{p\,j+2}+\cdots+\delta_{p\,i}\right)\Delta_{i\,j}^{x,y,z} \tag{3.24}$$

and we have introduced the usual Heaviside function defined by

$$\Theta(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0, \end{cases}$$

which arises due to the fact the summations in equation (3.21) are dependent on index i.

Each application of Algorithm 1 requires a user to specify the triplet of parameters (n_1, n_2, k_0), where k_0 is the spring constant of the harmonic potential used for $\gamma =$

Algorithm 2 *The NHCM with a determinant based re-looping condition*

Solve $\mathbf{F} = F(\mathbf{q})$ for $\gamma = 0$ by using the Gaussian elimination. The resulting configuration is denoted as $\mathbf{\Delta}^{0,n_2+1} \in \mathbb{R}^{3N-3}$.
Set $\Delta\gamma = 1/n_1$.
for $i = 1, 2, \ldots, n_1$ **do**
  Set $\gamma = i\,\Delta\gamma$ and $\mathbf{\Delta}^{i,1} = \mathbf{\Delta}^{i-1,n_2+1} \in \mathbb{R}^{3N-3}$.
  for $j = 1, 2, \ldots, n_2$ **do**
    Use equation (3.22) to calculate configuration $\mathbf{\Delta}^{i,j+1}$.
    Calculate forces corresponding to configuration $\mathbf{\Delta}^{i,j+1}$ to check convergence to $\mathbf{F}$.
  end for
  If convergence has been achieved during i and $i-1$ Newton's methods, and the determinant has decreased between the two: start the process from a previous iteration that did not converge ($j < i-1$) by setting $i = j-1$ and $\mathbf{\Delta}^{j-1,n_2+1} = \mathbf{\Delta}^{j,n_2+1}$
end for

0. The appropriate choice of these parameters is discussed in Section 3.4.1, where Algorithm 1 is applied to an LJ fluid. Algorithm 1 can also be modified by including a suitable stopping criterion for each Newton's method, rather than using n_2 iterations in each step. This is done in Algorithm 2 by testing the convergence for each Newton's method (at $\gamma = i\Delta\gamma$) and keeping track of the magnitude of the determinant of Jacobian matrix J: if the algorithm is converging for every γ but the determinant is steadily decreasing, we find looping back to a previous Newton's method ($\gamma = j\Delta\gamma$, where $j < i$) and running the method for another run of n_2 loops can sometimes solve this issue (as shown in figure 3.6 (a)). We present the findings in Section 3.4.1 where we discuss how the triplet (n_1, n_2, k_0) can be chosen to successfully invert given target forces $\mathbf{F}$.

3.4 Calculation of mapping (1.9) for an LJ fluid

As our illustrative example, we apply the NHCM to the LJ potential

$$U(r_{ij}) = 4\left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^{6}}\right), \tag{3.25}$$

where $k(r)$ in equation (3.17) is given by

$$k(r) = 24\left(\frac{2}{r^{14}} - \frac{1}{r^{8}}\right) \tag{3.26}$$

and factors (3.24) are given by

$$a_{ij} = \left(\frac{-48 \times 14}{|\Delta_{i\,j}|^{16}} + \frac{24 \times 8}{|\Delta_{i\,j}|^{10}}\right).$$

The LJ potential (3.25) tends to infinity as $r_{ij} \to 0$. In particular, if we used a uniform distribution of atoms as in example in Figure 3.1, we could obtain arbitrarily large (unphysical) forces. In principle, the following estimation of how large k_0 should be would still work, though in practice one would not typically see these large forces corresponding to unphysically close distances. In fact, if one were to have an incredibly dense configuration of atoms giving rise to large forces, then spreading the system out on a lattice corresponding to transformation (3.11) would potentially be unnecessary as the harmonic inversion gives rise to dense configurations as is.

Therefore, in the illustrative simulations, we use a randomly perturbed lattice configuration as illustrated in Figure 3.2(a), where the configuration $\mathbf{q}$ is given by

$$\mathbf{q} = \mathbf{q}^0 + \mathbf{q}^r. \tag{3.27}$$

Here, $\mathbf{q}^0$ is given by the regular cubic lattice (3.10) and the random component is $\mathbf{q}^r = (\mathbf{q}_1^r, \mathbf{q}_2^r, \ldots, \mathbf{q}_N^r)$ with $\mathbf{q}_i^r$ uniformly sampled in domain

$$\left[-\frac{\alpha\, L_x}{2\, N_x}, \frac{\alpha\, L_x}{2\, N_x}\right] \times \left[-\frac{\alpha\, L_y}{2\, N_y}, \frac{\alpha\, L_y}{2\, N_y}\right] \times \left[-\frac{\alpha\, L_z}{2\, N_z}, \frac{\alpha\, L_z}{2\, N_z}\right].$$

The configuration (3.27) depends on parameter $\alpha \in [0, 1]$ with $\alpha = 0$ corresponding to the regular lattice (3.10) and $\alpha = 1$ corresponding to atom positions uniformly

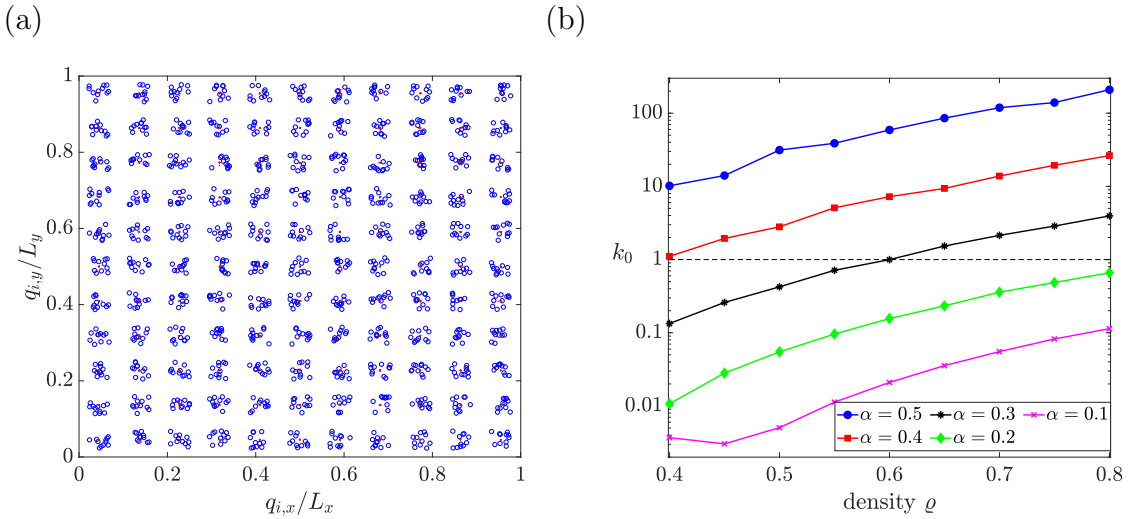


Figure 3.2: (a) *The positions* $\mathbf{q}_i$, $i = 1, 2, \ldots, N$, *of* $N = 1331 = 11^3$ *atoms which have been sampled in the cuboid domain* Ω *using* (3.27) *for* $\alpha = 0.5$ *and* $\varrho = 0.7$*(blue circles). We plot their scaled coordinates* $q_{i,x}/L_x$ *and* $q_{i,y}/L_y$ *in the* x*-*y *plane. The atoms positions in configuration* $\mathbf{q}$ *given by* (3.11)*, which solves equation* $\mathbf{F} = F(\mathbf{q})$ *for the harmonic potential are plotted as red dots.*
(b) *The value of the parameter* k_0 *given by equation* (3.29) *for density* ϱ *in the interval* $[0.4, 0.8]$ *and for different values of parameter* α *used in* (3.27)*. The black dashed line corresponds to* $k_0 = 1$*. This value of* k_0 *corresponds to how strong we must make the spring constant in order to achieve given forces.* k_0 *increases given the density* (ϱ) *or the lattice shuffling* (α) *increases as the forces necessarily become larger for the LJ potential.*

sampled in the whole domain Ω. Given $\alpha \in [0, 1]$, the sampled positions satisfy

$$r_{ij} \geq (1 - \alpha) \min \left\{ \frac{L_x}{N_x}, \frac{L_y}{N_y}, \frac{L_z}{N_z} \right\}.$$

Consequently, all pairwise forces are bounded for any $\alpha < 1$. In Figure 3.2(a), we use $\alpha = 0.5$ and $N = 1331 = 11^3$, showing that the atoms (blue circles) are well separated. Note that we only plot projections into x-y in Figure 3.2(a), i.e. the 11 atoms around each lattice point are separated in the z-direction. Considering $N = 11^3$ fixed, we can adjust the density of the LJ fluid, ϱ, by varying L_x, L_y, L_z. We use

$$L_x = L_y = L_z = \left(\frac{N}{\varrho} \right)^{1/3}. \tag{3.28}$$

With this choice, we can calculate the force vector $\mathbf{F} \in \mathbb{R}^{3N}$ for any given density ϱ.

The starting configuration for the NHCM is obtained for $\gamma = 0$ by using (3.11). It is plotted using red dots in Figure 3.2(a). To calculate the red dots, we use (3.28) with density $\varrho = 0.7$. The LJ potential (3.25) is used to calculate the force vector $\mathbf{F} \in \mathbb{R}^{3N}$ corresponding to the configuration plotted in Figure 3.2(a) as blue circles. In order to use (3.11) to calculate these red dots, we need to make sure that $\mathbf{F} \in \mathcal{F}$, where $\mathcal{F}$ is given by formula (3.8), and that the configuration $\mathbf{q}$ given by equation (3.11) solves the equation $\mathbf{F} = F(\mathbf{q})$. To achieve this, we choose k_0 large enough so that the sufficient condition (3.12) holds. Using (3.28), we can rewrite the condition (3.12) as

$$\| \mathbf{F}_i - \mathbf{F}_j \|_\infty \leq \frac{Nk}{2\,\varrho^{1/3}} \qquad \text{for all } i = 1, 2, \ldots, N, j = 1, 2, \ldots, N,$$

where $\| \cdot \|_\infty$ is the standard (unscaled) L^∞-norm. Therefore (3.12) will be satisfied if we choose k_0 as

$$k_0 \;=\; \frac{2\,\varrho^{1/3}}{N} \max_{i,j=1,2,\ldots,N} \| \mathbf{F}_i - \mathbf{F}_j \|_\infty \tag{3.29}$$

giving $k_0 = 144.2$ for the illustrative configuration in Figure 3.2(a). In Figure 3.2(b), we calculate k_0 using (3.29) for different values of density ϱ and the noise parameter α. Each point is calculated as an average over 10 realizations. We observe that k_0 has to be chosen larger for larger densities and for larger values of α.

3.4.1 Convergence of NHCM approach for an LJ fluid

The convergence of proposed Algorithms 1 and 2 are discussed in this section with $\Omega = \mathbb{R}^{3N}$. Both share an initial input of (n_1, n_2, k_0), so we first investigate the issues of pragmatically choosing these parameters.

The most common issue faced is that we would like to choose parameters n_1 and n_2 as small as possible, but without affecting the precision of the solution obtained due to the fact that the computational complexity increases linearly with both n_1 and

n_2. Choosing n_1 too small will result in $\Delta\gamma$ being too large, so when we input the last configuration as the starting configuration for the next Newton's method: the systems are too disparate and the ending configuration is no longer a sufficiently good starting guess. This is predominantly due to insufficient relaxing of the system: that is, the potential in (3.15) is dominated by either the harmonic or LJ term despite the co-factors of each monitoring this.

For this reason, we typically pick n_1 such that $\Delta\gamma \sim o(0.1)$. This can be seen in Figure 3.3 where convergence is seen for each system in this regime.

If n_2 is insufficiently large, a Newton's method does not have enough time to converge, often ending on a configuration which is not an appropriate initial guess for starting the subsequent Newton's method. This is not to say that all methods need to converge, indeed for $\gamma = i\Delta\gamma$ with i small, we rarely see convergence but for larger i, convergence is typically achieved. The goal is for convergence to be achieved when $i = n_1$ (or $\gamma = 1$). This is exemplified in Figure 3.3 where we see the bands of divergent parameter space when n_2 is small, regardless of the choice of n_1.

To demonstrate the problem faced by this general NHCM approach for force inversion, a system of $N = 5$ atoms with random forces $\mathbf{F} \in \mathcal{F}$, created from a normally distributed psuedorandom number generator, reveals a common theme. When the correct parameters are chosen, both algorithms will work in $n_1 \times n_2 + 1$ steps (we have n_1 Newton's methods and each runs for n_2 iterations with the additional first harmonic inversion which takes a single step). We see that convergence occurs (where the determinant flattens out in Figure 3.4) until the next Newton's method is initiated. Well chosen parameters demonstrate an eventual convergence of the determinant upwards (Figure 3.4(a)), whereas ill chosen parameters incrementally step the determinant down to 0 (Figure 3.4(c)). This leads to divergence in the final Newton's method when $\gamma = 1$.

There is still convergence in individual Newton's methods with $\gamma = i\Delta\gamma$ for the

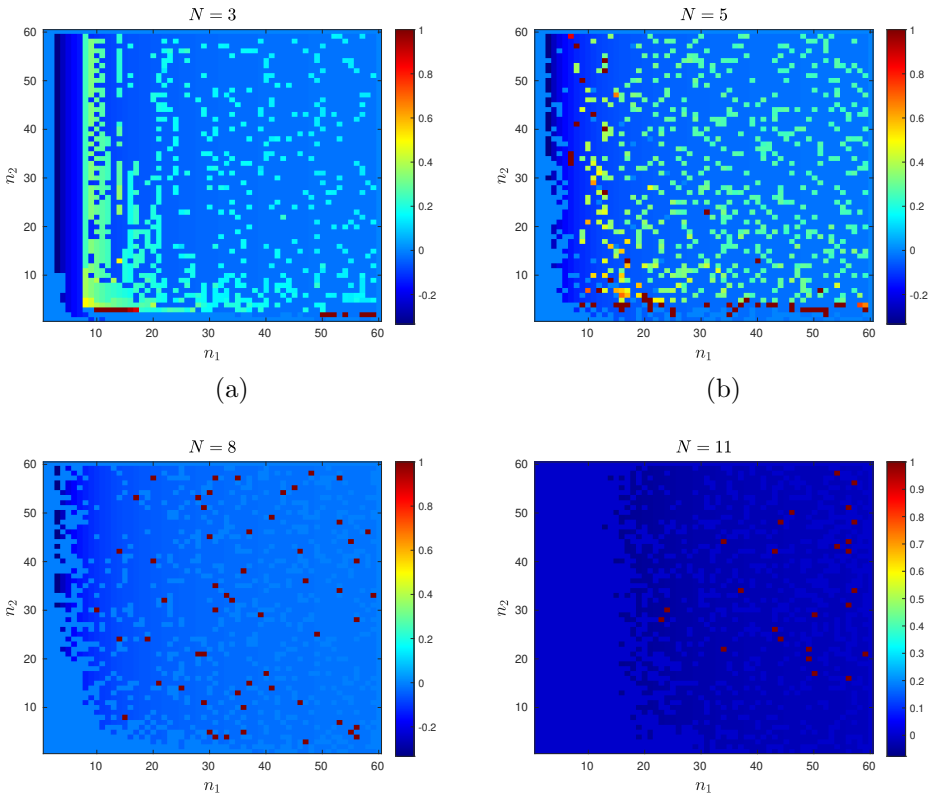


Figure 3.3: *In this figure, we illustrate the behaviour of parameter space* (n_1, n_2) *for Algorithm 1 with* $n_1 \in [1, 59]$ *and* $n_2 \in [1, 59]$ *for* $N = 3, 5, 8, 11$ *atoms and* $k_0 = 0.5$. *Whilst probing the parameter space, a value of* $z \in [-1, 1]$ *characterises how well the parameters* (n_1, n_2) *have performed. The value of* $|z|$ *is obtained as the first* γ *where convergence have been achieved (based on if the determinant between each iteration of* n_2 *for a given* γ *is less than a tolerance defined by* $tol = 10^{-10}$*). The sign is determined as positive if the method finally converges for* $\gamma = 1$, *and negative for divergence when* $\gamma = 1$. *Any parameter pair* (n_1, n_2) *where convergence is not achieved within the tolerance, but the algorithm has converged to the target forces at* $\gamma = 1$, *is automatically given the value of* 1. *We see the parameters become much harder to find as the configuration size increases.*

NHCM with parameter choices that lead to divergence: where i is close but not equal to n_1. These are explained as having converged to a bad configuration early on in the inversion, and the convergence per step is due to the fact there is a non-zero harmonic term that prevents a non-singular Jacobian at this configuration. However once we move to the last Newton's iteration with $i = n_1$ (when $\gamma = 1$), this harmonic perturbation is removed and evaluation of the inverse Jacobian is now impossible.

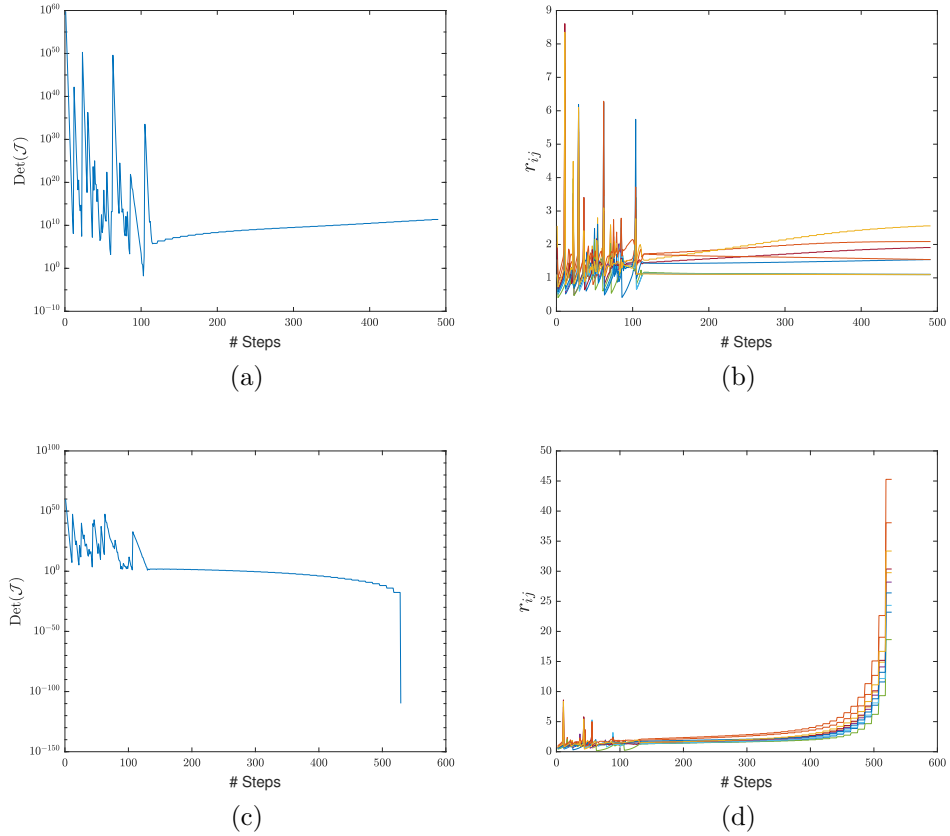


Figure 3.4: *Panels (a) and (b) correspond to running Algorithm 1 with* $(n_1, n_2) = (49, 9)$*, whilst panels (c) and (d) correspond to* $(n_1, n_2) = (50, 9)$*, all for* $N = 5$*. Panels (a) and (b) typify wanted convergent behaviour as the determinant steadily increases until the end of the inversion, whilst the* $N(N-1)/2$ *pairwise distances* r_{ij} *also converge. This contrasts the ill chosen parameters in (c) and (d) where the determinant eventually converges for each Newton's method, though it steadily decreases to* 0*, which results in the pairwise distances diverging (see equation (3.22)) as the harmonic forces dominate and try to hold the system together. We omit the final* n_2 *steps in panel (d) due to the divergent nature of distances.*

More specifically, with a divergent parameter choice of (n_1, n_2), convergence for subsequent Newton's methods can be achieved, but the interatomic distances begin to grow. This is a direct cause of the fact that harmonic forces begin to dominate each step. Overcoming the retarding factor of $(1-\gamma)$ necessarily means the harmonic force has to increase which is done by increasing the interatomic distances. This becomes more exaggerated as the LJ forces fall off throughout the inversion. On the last Newton's method when $\gamma = 1$ and the harmonic forces are turned off, the

systems are too different for any hope of convergence which leads to numerical errors. We see in Figure 3.5(b) that the harmonic forces for this set of parameters for $N = 5$: $(n_1, n_2) = (49, 10)$, do not converge to zero before they are turned off, and this discontinuity gives divergence in the final Newton's method.

The questions that arise are: how can one tell if a configuration is 'bad', and how can we fix this for each choice of parameter values (n_1, n_2)? In answer to this first question: it is still not clear what typifies a 'bad' configuration at the time it is produced, except the eventual decreasing convergent behaviour of the determinant, and convergence of the harmonic forces to a non-zero force before it is switched off. We use these as conditions to rectify the force inversion when a bad parameter set is chosen.

The second question is of more practical value; we have probed small systems of atoms ($N < 10$) where convergence is typically determined after the first few Newton's methods are employed. Given parameter n_1, we can see there are many n_2 which correspond to converging force inversions in Figure 3.3. We have taken configurations $\mathbf{\Delta}^{i,n_2+1}$ from successful choices of n_2 and used them as the starting configuration (at the required truncation) $\mathbf{\Delta}^{i+1,1} = \mathbf{\Delta}^{i,n_2+1}$ for a previously divergent n_2. This successfully remedies the divergent issues for a given (n_1, n_2), though relies on *a priori* knowing a convergent set of parameters and slotting in one of the configurations.

This can be seen in Figure 3.5(b) where the 'good' configuration from $N = 5$ with $(n_1, n_2) = (49, 9)$ is taken after the 10^{th} Newton's method, $\mathbf{\Delta}^{10,9+1}_{\text{good}}$, and imported into the $(n_1, n_2) = (49, 10)$ divergent inversion where $\mathbf{\Delta}^{11,1}_{\text{bad}} = \mathbf{\Delta}^{10,9+1}_{\text{good}}$. The harmonic forces converge to zero and this force inversion is now successful.

One way to implement a fix more autonomously is to loop the simulation again when either of these 'bad' behaviours are eventually seen. For small systems, re-looping the first Newton's method for extra iterations is successful because this is make or break for the system, so to speak. However for a larger system (typically

$N > 4$), these can break down during subsequent Newton's methods. A potential fix would be to randomly revisit a Newton's method before convergence to the target force is achieved and re-loop for a random number of additional iterations. This method is not recommended for practical reasons as there is no bound on how long it will take to eliminate the divergent behaviour because there is currently no defining criterion for a 'good' configuration at the time of creation.

However for a small system with $N = 3$ atoms, we implement the re-looping condition where the determinant converging in a decreasing manner is identified and the system is made to run for another n_2 iterations on the first Newton's method (when $n_1 = 1$). Figure 3.6(a) demonstrates that the parameter space is now more accessible compared to Figure 3.3(a). We introduce a tally that denotes the maximum number of times we re-loop for an additional n_2 steps before we search for a different (n_1, n_2) parameters.

(a) (b)

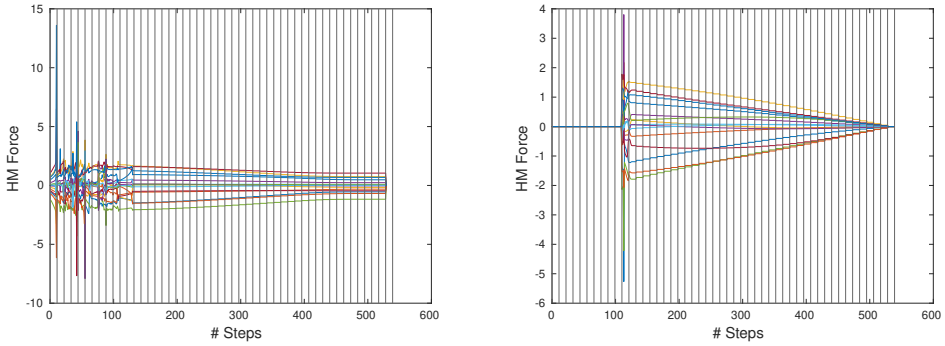
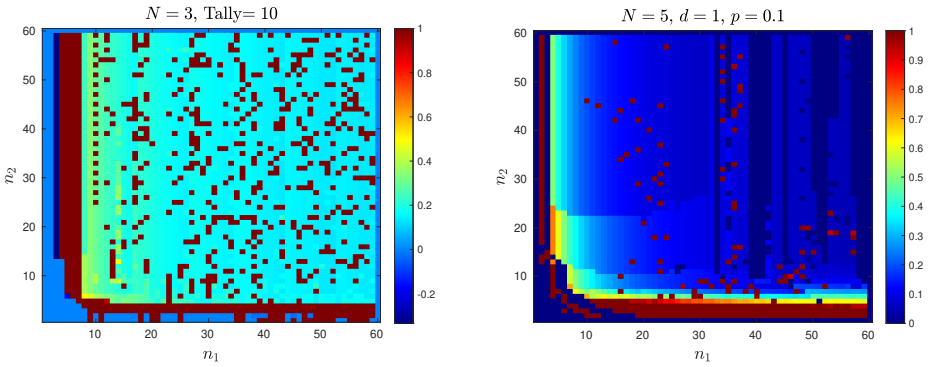


Figure 3.5: *With a system of $N = 5$ atoms an inversion with $n_1 = 49$, $n_2 = 10$ is divergent which is typified by the incorrect harmonic force convergence shown in panel (a). However when a good configuration from $n_1 = 49$, $n_2 = 9$ is placed into this previously ill behaved force inversion after the 10^{th} Newton's method, we see the correct convergent behaviour illustrated in panel (b). This indicates that a possible modification is to re-loop from random previous steps before any convergence is achieved. The vertical black lines in this figure partition each Newton's method corresponding to n_2 steps.*

A more conducive way to open up the parameter space of successful choices for (n_1, n_2) is to bias the system. To reduce the harmonic force domination, we introduce

a truncation of any harmonic term with a factor of $\exp(-r_{ij}^2)$ when $r_{ij} > d$, where d is a truncation parameter. Here, larger distances r_{ij} penalise the harmonic force in an attempt to stop it growing. This alone helps to keep the distances from growing too large though the harmonic forces can still converge to a non-zero force before they are switched off as is the case in Figure 3.5(a). To address this an additional tuning factor, $k_0 \to k_0 \exp\left(1 - \frac{1}{(1-\gamma)^p}\right)$, is introduced where p determines how severely we wish to reduce the harmonic forces.

Although one may think that adding new parameters may complicate the problem, in practice these are much easier to toggle and one can find appropriate (d, p) for each N (given a target force). This increases the likelihood that any random parameter choice (n_1, n_2) does converge. This is illustrated in Figure 3.6(b) where $N = 5$. In comparison to Figure 3.3(b) we now see that almost the whole parameter space exhibited for $(d, p) = (1, 0.1)$ successfully inverts the target force $\mathbf{F}$.



(a) (b)

Figure 3.6: *Panel (a) illustrates the use of Algorithm 2 with a re-looping condition with a maximum re-looping of* 10 *times. This can be compared with Figure* 3.3 *(a), we see the accessible parameter space has opened up significantly. Similarly panel (b) indicates the use of tuning parameters* $(d, p) = (1, 0.1)$ *for the* $N = 5$ *system of atoms, compared with Figure* 3.3 *(b), again the* (n_1, n_2) *parameter space has been opened as seen by a higher density of successful choices.*

It is worth noting the physical interpretation of the system at play: we have introduced a harmonic potential with an equilibrium length of $r^* = 0$, which is purely

attractive, whereas the LJ potential has repulsive and attractive regimes which admit an equilibrium length of $r^* = 2^{1/6}$. It is natural to assume that toggling parameters in systems with different characteristics is not the best way to proceed. However, the analytical tractability of this unshifted harmonic potential is a benefit that outweighs other factors. One can use a harmonic potential with an identical equilibrium length of $r^* = 2^{1/6}$ as

$$U_0(r_{ij}) = \frac{k_0}{2}\Big(r_{ij} - r^*\Big)^2. \tag{3.30}$$

With the shifting incurred by equation (3.30), two further terms are added to equation (3.23), which is modified to include summation over the harmonic terms as $k_0 \to k_0(r_{ij} - r^*)/r_{ij}$ (with $i > j$):

$$\begin{aligned}\frac{\partial \mathbf{H}_i}{\partial \mathbf{\Delta}_{p\ p-1}} = (3.23) &+ \sum_{\ell=2}^{i}\left[\Delta_{\ell\,\ell-1}\left[\sum_{j=1}^{\ell-1} k_0 \frac{(1-\gamma)r^*}{r_{ij}^3}(\delta_{p\,j+1} + \delta_{p\,j+2} + \cdots + \delta_{p\,i})\Delta_{i\,j}\right]\right] \\ &- \sum_{\ell=i+1}^{N}\left[\Delta_{\ell\,\ell-1}\left[\sum_{j=1}^{\ell-1} k_0 \frac{(1-\gamma)r^*}{r_{j\,i}^3}(\delta_{p\,i+1} + \delta_{p\,i+2} + \cdots + \delta_{p\,j})\Delta_{j\,i}\right]\right].\end{aligned} \tag{3.31}$$

This has the benefit of reducing occurrences in the force inversion where interatomic distances get closer to zero (by the harmonic force). The LJ forces then take over and repel atoms away from each other (which is seen in Figure 3.4(b)). However we have found the final step of these force inversions to be consistently divergent despite the re-looping and truncation conditions. Practically the un-shifted harmonic potential with truncation conditions does a better job of stopping distances from getting non-physically large, reducing the effect of the harmonic forces both taking over and converging to a non-zero force before the final Newton's method.

Finally a note on choosing the spring constants: in this Section, we present results where we used $k_0 = 0.5$ though the analysis carries through for general k_{ij}. One could potentially create a NHCM for each pairwise interaction indexed by i, j and turn each

interaction from harmonic to LJ one by one, though this would scale with at least an additional factor of $O(N)$. We have found that choosing spring constants that are too small does not allow the system to relax into the final configuration, as the first inversion creates a dense configuration that is ill behaved when short range repulsive potentials such as LJ are use. We cannot use the shifting introduced in the PBC case for the general setting. It is therefore harder to swap from the harmonic case to the LJ case smoothly.

3.5 Force inversion for $N = 2$ atom systems

The $N = 2$ system is studied for pedagogical purposes: in this system we have two atoms with positions $\mathbf{q}_1$, $\mathbf{q}_2 \in \mathbb{R}^3$. Here we have a set of equations

$$F_1^x = -k(r_{12})\,\Delta_{21}^x \tag{3.32a}$$

$$F_1^y = -k(r_{12})\,\Delta_{21}^y \tag{3.32b}$$

$$F_1^z = -k(r_{12})\,\Delta_{21}^z \tag{3.32c}$$

where $F_1^{x,y,z}$ denotes each component of the force on atom 1. By Newton's third law, the force on atom 2 is trivially $\mathbf{F}_2 = -\mathbf{F}_1$. We see the system of equations $\mathbf{F}_1 + k(r_{12})\Delta_{21} = \mathbf{0}$ is already written in terms of single jump displacements. Given $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2]$: we aim to solve equations (3.32) for Δ_{21}. With this, we can reconstruct the positions $\mathbf{q}_1$ and $\mathbf{q}_2$ up to a translation in space.

Given existence, we can prove uniqueness of the solution Δ_{21}^* in this 2-atom case, subject to (3.33), by showing that the function $\mathbf{H}_1(\mathbf{F_1}, \Delta_{21}^*) = \mathbf{0}$ is locally invertible. Writing equations (3.32) as

$$\mathbf{H}_1 = \mathbf{F}_1 + k(r_{12})\,\Delta_{21},$$

we can compute the Jacobian as

$$J(\mathbf{H}_1)_{ij} = \frac{\partial\big(\mathbf{H}_1\big)_i}{\partial\big(\Delta_{21}\big)_j} = a_{12}\,\Delta_{21}^j\,\Delta_{21}^i + k(r_{12})\,\delta_{ij},$$

where δ_{ij} is the Kronecker delta function and prefactors a_{ij} are determined by

$$\frac{\partial k(r_{12})}{\partial\big(\Delta_{21}\big)_j} = a_{12}\,\Delta_{21}^j,$$

where $i, j = 1, 2, 3$.

The determinant of J evaluated at Δ_{21} is given by

$$\det(J) = a_{12}\,k(r_{12})^2\left[|\Delta_{21}|^2 + \frac{k(r_{12})}{a_{12}}\right].$$

A condition for local invertibility is $\det(J) \neq 0$, which reduces to

$$|\Delta_{21}|^2 + \frac{k_{12}}{a_{12}} \neq 0. \tag{3.33}$$

In particular the LJ potential satisfies the condition (3.33) locally, given that $r_{12} = r_{21} \neq (26/7)^{1/6}$. The 2-atom system with a purely LJ potential can be solved with the added intricacy that there are three branches on which solutions can be found. This arises from the fact that when tackling equations (3.32) with prefactor $k(r_{12})$ in (3.26) one can form the following expressions (noting that $k(r_{12}) \neq 0$):

$$\Delta_{21}^y = \frac{F_1^y}{F_1^x}\,\Delta_{21}^x \qquad \text{and} \qquad \Delta_{21}^z = \frac{F_1^z}{F_1^x}\,\Delta_{21}^x. \tag{3.34}$$

Noting that $|\Delta_{21}| = \sqrt{(\Delta_{21}^x)^2 + (\Delta_{21}^y)^2 + (\Delta_{21}^z)^2} = r_{12}$, we find

$$r_{12}^2 = (\Delta_{21}^x)^2\left(\frac{|\mathbf{F}_1|^2}{({F_1}^x)^2}\right),$$

which we then substitute into equation (3.32a) resulting in the polynomial

Regions in the Lennard-Jones force-displacement plane			
Region 1	Region 2	Region 3	Region 4
$F(x) < -2.396$	$-2.396 < F(x) < 0$	$0 < F(x) < -2.396$	$F(x) > 2.396$
$-1.0913 < x < 0$	$x > 1.2445$	$1.0913 < x < 1.122$	$0 < x < 1.0913$
	$1.122 < x < 1.2445$	$-1.2445 < x < -1.122$	
	$-1.122 < x < -1.0913$	$x < -1.2445$	

Table 3.2: *Here we categorise possible roots in four distinct regions in the LJ force-displacement plane illustrated in Figure* 3.7.

$$F_1^x \left(\Delta_{21}^x\right)^{13} + \frac{24(F_1^x)^8}{|\mathbf{F}_1|^8} \left(\Delta_{21}^x\right)^6 - \frac{48(F_1^x)^{14}}{|\mathbf{F}_1|^{14}} = 0. \tag{3.35}$$

Upon solving equation (3.35) for Δ_{21}^x, we can find the orthogonal y and z components using expressions (3.34), thus finding all components of $\Delta_{21} \in \mathbb{R}^3$.

The multiple branches of solutions are inherited from equation (3.35), or originally from (3.16) which in one dimension (with position $x \in \mathbb{R}$) breaks down into

$$F(x) = \frac{48}{x^{13}} - \frac{24}{x^7}.$$

We see from Figure 3.7 that the force can fall into four distinct regions where the LJ potential admits a local minimum and maximum of $F(x) = \pm 2.396\ldots$ which separate the one and three solutions regions. Roots in these regions are categorised in Table 3.2.

Regarding the initial problem in this subsection, if we have a total force $|\mathbf{F}| > 2.396\ldots$ then equation (3.35) will only have one solution, giving a unique Δ_{21}, since the total force falls in region 4. However if the total force $0 < |\mathbf{F}| < 2.396\ldots$ falls in region 3, then we will find three solutions for Δ_{21}.

One can make sense of this intuitively: if an atom is experiencing a large force, there is a short range Pauli repulsion which causes unboundedly large forces. On the other hand, weaker forces can be caused be a repulsive force from one side, or attractive forces drawing the atom from the opposite side. The LJ potential admits

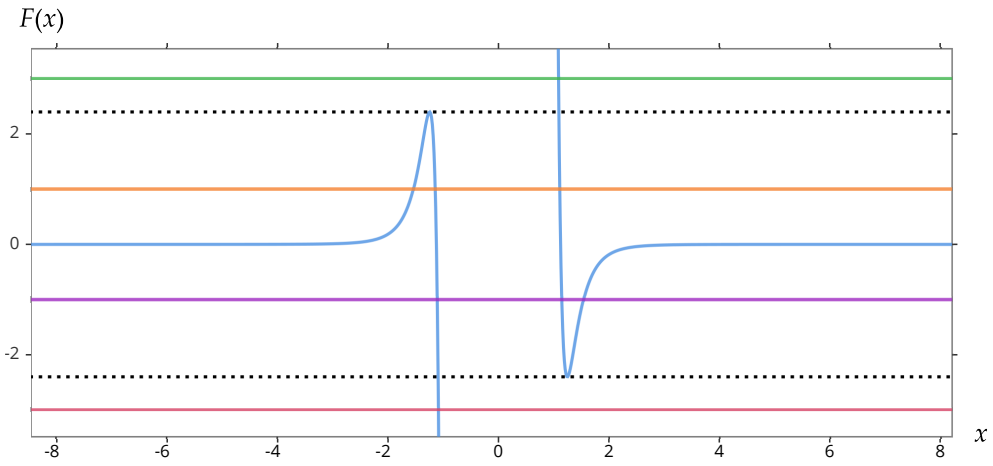


Figure 3.7: *In this illustration, we show that when we have* $F(x) = \pm 3$*, there is a unique solution which is indicative that we are in regions* 1 *or* 4 *respectively. In regions* 2 *and* 3*, there are three solutions (illustrated here when* $F(x) = \pm 1$*). The dotted lines denote the barriers* $F(x) = \pm 2.396\ldots$ *which separate the disparate regions. For example, the green line corresponds to a force above* $F(x) = 2.396\ldots$*, there is only one distance corresponding to this force due to short range repulsion which is illustrated in Figure 3.8. The orange line lies below this force barrier, and there are three distances that correspond to the given force, one short range repulsion and two attractive configurations, each also illustrate in Figure 3.8.*

two distinct positions from where an atom is attracted with the same force: hence there are three solutions in this case. These are exhibited in 1d in Figure 3.8.

We now present an algorithm inverting the forces obtained from MD simulation of two atoms in 3d, interacting via a LJ potential with PBCs. We also discuss how to navigate the intricacies involved with three potential configurations per force inversion. Importantly, we only use the data of the forces at each time step to produce the resulting configuration, which can be compared with MD position data stored for bench-marking purposes.

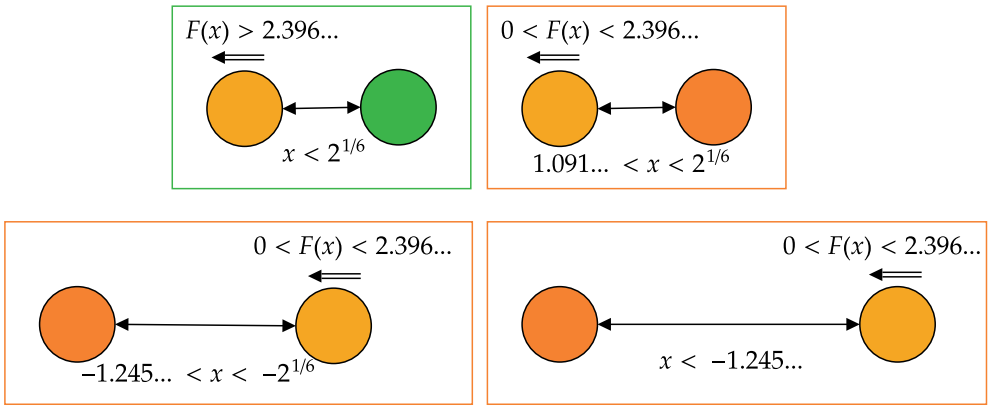


Figure 3.8: *The top left green panel corresponds to a force on atom* 1 *(the orange atom) in region* 4 *where there is only one solution, a short range Pauli repulsion, that produces the required force. A force found in region* 3 *has three possible configurations, one corresponding to short range Pauli repulsion (top right) and two from an attractive interaction pulling on the opposite side of atom* 1 *(bottom left and right). The green/orange box(es) corresponds to the green/orange lines in Figure 3.7.*

We define the force data for n_t time steps as

$$\underline{\underline{\mathbf{F}}} = \begin{pmatrix} F_1^x(1) & F_1^y(1) & F_1^z(1) \\ F_1^x(2) & F_1^y(2) & F_1^z(2) \\ \vdots & \vdots & \vdots \\ F_1^x(n) & F_1^y(n) & F_1^z(n) \end{pmatrix} \in \mathbb{R}^{n_t \times 3},$$

where one can invert each row of $\underline{\underline{\mathbf{F}}}$ with equations (3.34,3.35) to reconstruct Δ_{21}.

As discussed previously, there is at least 1 and at most 3 solutions to (3.35). We calculate the resulting configurations and distances

$$rd_i(j) = \sqrt{(\Delta_{21}^x(j))^2 + (\Delta_{21}^y(j))^2 + (\Delta_{21}^z(j))^2},$$

where j indexes the time step $1 \leq j \leq n$ and $i = 1, 2, 3$ corresponds to the possible multiple solutions.

These are ordered such that $rd_3(j)$ always corresponds to short range repulsion and $rd_1 \geq rd_2$. If a force at time step $j = 1, \ldots, n$ satisfies $|\mathbf{F}_1(j)| > 2.396\ldots$, we

are necessarily in region 1 or 4 of Figure 3.7. In this case, we set $rd_1 = rd_2 = 0$. This does not represent a configuration with zero distance and is merely set for convenience while storing data. Once these distances have been calculated for each time step, we can construct the vectors $\mathbf{rd}_1, \mathbf{rd}_2, \mathbf{rd}_3 \in \mathbb{R}^{n_t}$.

The problem arises when piecing together the configurations in a coherent way that represents physical dynamics: for example, we cannot arbitrarily switch from $rd_2(j)$ to $rd_1(j+1)$ to $rd_3(j+2)$ in three subsequent time steps. The aim is to construct $\mathbf{r} \in \mathbb{R}^{n_t}$ where each component represents the distance between the two atoms (namely r_{12}) at a given time step.

If the initial configuration is known, we can calculate the distance $r(1)$ and match it with $rd_1(1), rd_2(1), rd_3(1)$ to start the reconstruction process. If the initial distance is not known, we choose one of the potential roots to start with and only one will give rise to physical dynamics. In the system we studied, this was rd_1 which corresponds to long range attraction. This is the most likely choice when you have a low density system.

Next, it is important to understand that $rd_3 < 1.122\ldots$, $1.122\cdots < rd_2 < 1.244\ldots$ (or $rd_2 = 0$), and $1.244\cdots < rd_1$ (or $rd_1 = 0$). The roots in regions 1-4 are given in Table 3.2 and we can see there are barriers at $r = 1.122\ldots$ and $r = 1.244$. We posit that if the 'track' of rd_1,rd_2 or rd_3 crosses one of these barriers, we must switch to an adjacent 'track'. We list and illustrate the possibilities that arise when reconstructing $\mathbf{r} = (r(1), r(2), ..., r(n)) \in \mathbb{R}^{n_t}$ in Table 3.3 and Figure 3.9 respectively.

Lastly, we need to automate the identification of these switches, and we can do so by looking at previous data and using some dynamical information. For example, in an ID switch, we are necessarily on the rd_1 track and trying to identify when we switch on to a rd_2 track. We do so by looking at the sequence $rd_1(j-2)$, $rd_1(j-1)$, $rd_1(j)$ and $rd_1(j+1)$. An ID switch is formed by identifying a sequence structure such as $rd_1(j-2) > rd_1(j-1) > rd_1(j) < rd_1(j+1)$. We decrease until we reach the

Reconstructing $\mathbf{r}$ from time step j to $j+1$		
$r(j)$	$r(j+1)$	Type
$rd_1(j)$	$rd_1(j+1)$	Same track
$rd_2(j)$	$rd_2(j+1)$	Same track
$rd_3(j)$	$rd_3(j+1)$	Same track
$rd_1(j)$	$rd_2(j+1)$	ID
$rd_2(j)$	$rd_3(j+1)$	IID
$rd_3(j)$	$rd_2(j+1)$	IIU
$rd_2(j)$	$rd_1(j+1)$	IU

Table 3.3: *While reconstructing the vector of distances* $\mathbf{r}$*, we need to make appropriate choices while following the different roots of equation* (3.35) *in a consistent manner. This may require switching from the* rd_1 *track for time step* j *to the* rd_2 *track for time step* $j+1$*, which we call a ID switch.*

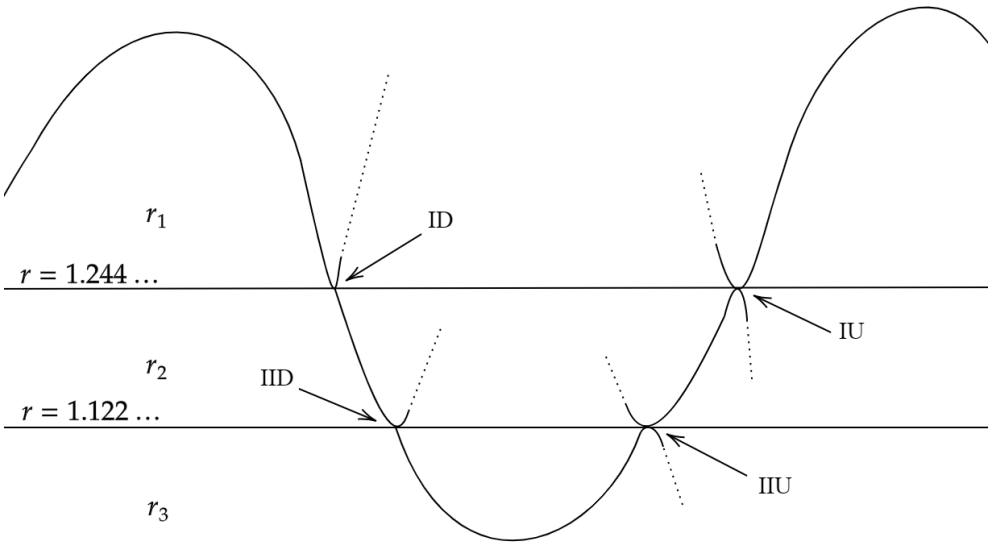


Figure 3.9: *As we follow the different distance tracks when reconstructing* $\mathbf{r}$*, we must make the choice to switch to the appropriate track when a barrier is hit: these are categorised as ID, IID, IIU and IU in Table* 3.3*.*

(a) (b)

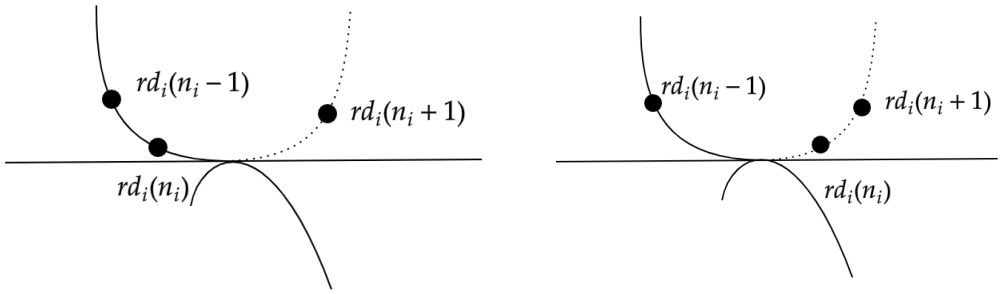


Figure 3.10: *In this figure we realise both possibilities after obtaining the necessary relation* $rd_i(n_i-1) > rd_i(n_i) < rd_i(n_i+1)$ *for an ID or IID switch. In case (a) the root switch happens from* n_i *to* n_i+1*: here* $r(n_i) = rd_i(n_i)$ *and* $r(n_i+1) = rd_{i+1}(n_i+1)$*. In case (b) the root switch happens from* $n_i - 1$ *to* n_i*: so* $r(n_i-1) = rd_i(n_i-1)$ *and* $r(n_i) = rd_{i+1}(n_i)$*. One can distinguish the two cases by looking at the subsequent jumps* $d_1 = |rd_i(n_i-2) - rd_i(n_i-1)|$ *and* $d_2 = |rd_i(n_i-1) - rd_i(n_i)|$*.*

barrier and 'bounce'. This indicates that we have met the barrier and must switch to a different track. However we need to understand at which time step this happens as illustrated in Figure 3.10.

In order to ascertain which of the two scenarios is most likely to happen: we look at the jump between $rd_1(j-2), rd_1(j-1)$ defined by $d_1 = |rd_1(j-2) - rd_1(j-1)|$ and the jump between $rd_1(j-1), rd_1(j)$ defined by $d_2 = |rd_1(j-1) - rd_1(j)|$. If the absolute value of the difference between jumps $d = |d_1 - d_2| < 0.001$ then it is extremely likely the sequence $rd_1(j-2), rd_1(j-1), rd_1(j)$ hasn't been altered by potentially dropping below the barrier and we are in the left panel of Figure 3.10. Conversely if this condition ($d = |d_1 - d_2| < 0.001$) is not achieved, it is likely that this disparity between jumps is due to a premature crossing of the barrier and we are in the alternative case (given by the right panel in Figure 3.10). The same method works to identify where an IID switch happens except we use rd_2 in place for rd_1 above.

The jump conditions characterise dynamical information because, given a sufficiently smooth potential, the forces should continuously change. These can be altered

based on the previous force jumps and the condition $d = |d_1 - d_2| < k$ can be varied for different constants k (which should be inferred from the previous distances). In the simulation we present, taking $k = 0.001$ for ID and IID switches, and $k = 0.005$ for IU and IIU switches, suffices to reproduce the configurations for each time step exactly. We utilise a very similar scheme for IU and IIU switches - except we look for a sequence of $rd_i(j-2) < rd_i(j-1) < rd_i(j) > rd_i(j+1)$ (where $i = 2, 3$ respectively).

Finally, we present Algorithm 3 which has been implemented on simulation data from two Argon atoms interacting via a LJ (12-6) potential with a Nosé-Hoover thermostat and PBCs. Here we have taken the temperature to be $T = 1$ and the box width to be $L_x = L_y = L_z = 4$. We choose to start with $r(1) = rd_1$ because in most of the simulation, the atoms are in a weakly attractive phase whilst the smaller roots rd_2 and rd_3 correspond to strong attraction and strong repulsion. These dynamics act relatively quickly, which can be seen by the fact that most components of $\mathbf{r}$ are made from rd_1. Sampling a random, consecutive $n_t = 1000$ time steps, we are able to store the force data and implement Algorithm 3, and compare against the actual position data from the MD simulation. Figure 3.11 shows that we are able to faithfully reproduce the distances between the atoms using this automated method which relies purely on force data.

3.6 Discussion and conclusions

In this chapter, we have presented the problem of force inversion and have provided an exploratory study addressing how one can approach this. The $N = 2$ system has been generalised in Section 3.5 with application to LJ forces which present additional challenges where a monotonic force (like gravity) wouldn't. Consideration is given to reconstructing a continuous choice of inversions during a dynamical procedure. Section 3.3 details the specifics of how a Newton homotopy continuation method can be used to switch from an analytically tractable system of harmonic atoms to that of

Algorithm 3 *Inverting a 2 atom system interacting via a LJ potential given force data* $\underline{\underline{\mathbf{F}}}$

Import MD force data as a matrix $\underline{\underline{\mathbf{F}}} = (\mathbf{F}(1), \mathbf{F}(2), \ldots, \mathbf{F}(n)) \in \mathbb{R}^{n_t x 3}$
for $j = 1, \ldots, n_t$ **do**
  if $|\mathbf{F}(j)| > 2.396\ldots$ **then**
    Solve equation (3.35) with $F_1^x = F_t^x(j)$, $F_1^y = F_t^y(j)$ and $F_1^z = F_t^z(j)$
    Use equations (3.34) to reconstruct $\Delta_{21}(j)$
    Calculate $rd_3(j) = \sqrt{(\Delta_{21}^x(j))^2 + (\Delta_{21}^y(j))^2 + (\Delta_{21}^z(j))^2} = |\Delta_{21}(j)|$
    Set $rd_1(j) = rd_2(j) = 0$
  else
    Solve equation (3.35) with $F_1^x = F_t^x(j)$, $F_1^y = F_t^y(j)$ and $F_1^z = F_t^z(j)$
    Use equations (3.34) to reconstruct three configurations $\Delta_{21}^i(j)$ where $i = 1, 2, 3$

    for $i = 1, 2, 3$ **do**
      Calculate $rd_i(j) = |\Delta_{21}^i(j)|$
    end for
    Sort such that $rd_1(j) > rd_2(j) > rd_3(j)$
  end if
end for
Check for ID switches
for $j = 3, \ldots, n_t$ **do**
  if $rd_1(j-2) > rd_1(j-1)$ **and** $rd_1(j-1) < rd_1(j)$ **then**
    Calculate $d_1 = |rd_1(j-2) - rd_1(j-1)|$
    Calculate $d_2 = |rd_1(j-1) - rd_1(j)|$
    Calculate $d = |d_1 - d_2|$
    if $d < 0.001$ **then**
      Identify $j+1$ as a ID switch
    else
      Identify j as a ID switch
    end if
  end if
end for
Similarly check for for IID, IIU and IU switches
Identify $r(1) = rd_1(1)$ (or choose another root to start the reconstruction)
for $j = 2, \ldots, n_t$ **do**
  $\# =$ number of switches up until (and including) j
  if $\# = 0 \mod 4$ **then**
    $r(j) = rd_1(j)$
  else if $\# = 1 \text{ or } 3 \mod 4$ **then**
    $r(j) = rd_2(j)$
  else if $\# = 2 \mod 4$ **then**
    $r(j) = rd_3(j)$
  end if
end for
for $j = 1, \ldots, n_t$ **do**
  Reproduce the configurations $\Delta_{21}(j)$ corresponding to the root identified by $r(j)$
end for

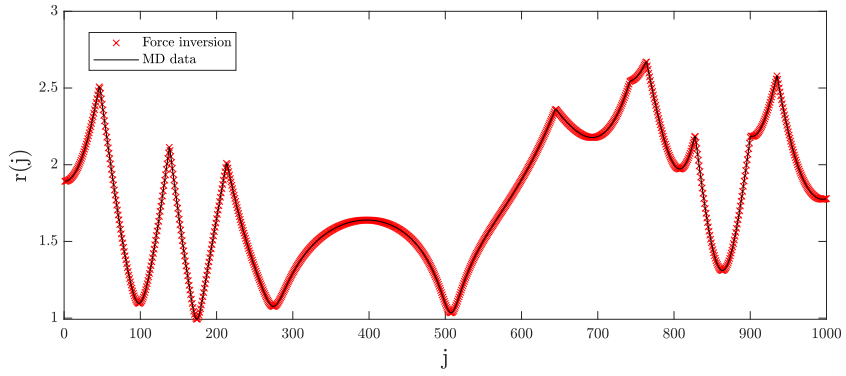


Figure 3.11: *Here we compare our force inversion algorithm against MD data and get accuracy up to floating point precision. We have demonstrated that the jump conditions with* $k = 0.001$ *for identifying ID and IID switches is appropriate here, and* $k = 0.005$ *for IIU and IU switches.*

a fully LJ system, whereby correct parameter choices for (n_1, n_2) can be found by re-looping conditions or toggling harmonic forces directly with truncation and stronger retardation factors.

It cannot be overstated how important developments in the study of N-atom systems are. These provide the basis for AA simulations in MD which parameterise the majority of semi-empirical bottom-up models [178]. These semi empirical methods include CG methods [186, 129] and mean field methods [176, 72], both ameliorate the computational complexity of the AA simulation, with the downside that dynamical accuracy is often lost [107].

A common criterion for parameterising CG models is force-matching. This allows the energy landscape of both systems to be as similar as possible with dynamics obtained as accurate as possible [125, 71] - this has been expounded in Section 1.3. The investigation in this chapter provides a theoretical insight into how configurations can be obtained from force data obtained from AA simulation. The force-matching approach has been introduced in Section 1.3 but more recently the use of machine learning has aided the development of this field with the invention of CGnets [70], where force data is used to train neural networks and gradients of predicted free

energy are used to estimate the PMF. Other ideas include incorporating relative entropy minimisation [138].

The aim of this work is to provide the foundation of answering the problem of force inversion, but also exhibiting and analysing a novel application of NHCM. In future we will focus on developing more rigorous criteria for identifying 'bad' configurations with an eye to probe larger N systems. A potential way to improve the scalability of the current method is to use machine learning to identify these convergent/divergent behaviours. Similarly modified NHCMs can be used with the aid of arc-continuation to potentially subvert the issues faced.

With this in mind, application to reproducibility of data within MD of simple fluids and force-matching CG are examples of areas that would benefit from force inversion. As well as directly reducing the complexity of systems with bonded interactions to harmonic systems. Many-body systems have been studied for centuries and have percolated through most areas of mathematics. Force inversion could play an important role where total forces can be measured for each atom or more generally object; even further afield to areas such as robotics (sensing).

Chapter 4

Symmetries of many-body systems imply distance-dependent potentials

The work in Chapter 4 is the result of research undertaken in the final year of my DPhil study. A subset of the work presented here has formed a paper [218] written in collaboration with Radek Erban, which has been accepted for publication in Physical Review E. The introduction has been merged with Section 1.6 and additional material is included in Section 4.6.

The aim of this chapter is to understand the classical underpinnings of non-reciprocal interactions introduced in Section 1.6, for which there seems to be a large, yet exciting gap to explore. The focus in this field has largely been application based, especially in the fields of Optics, Electromagnetism and Biochemistry. We begin this chapter with a theoretical consideration exploring the symmetries of a potential which obeys Newton's third law (both weak and strong forms).

Following this, we introduce the definition of a central potential, and the corresponding central potential theorem in Section 4.2. We also show that only a relatively

small subset of distances are required to uniquely determine the geometry of the system, and an even smaller subset of distances are needed to adequately evaluate the potential given the correct starting structure. These can then be used to describe the whole potential, which is illustrated in Section 4.6. General many-body potentials are considered and, similar to Richard et al [180], we develop the expression for the closed form of the potential energy, purely based on pairwise distances. Limitations of this distance description are further discussed and evaluated in Section 4.7.

4.1 Potentials for non-reciprocal interactions

As presented in Section 1.6, non-reciprocal interactions are pairwise interactions between atoms i,j whereby pairwise forces are governed by a two-body potential $U_2(\mathbf{q}_1, \mathbf{q}_2) \in \mathbb{R}$ with

$$\mathbf{F}_{ij} = -\nabla_i U_2 \in \mathbb{R}^3, \tag{4.1}$$

and $\mathbf{F}_{ij} \neq -\mathbf{F}_{ji}$.

Given a system of N atoms that obey Newton's third law (weak form), the sum of total forces is zero:

$$\sum_{i=1}^{N} \mathbf{F}_i = \sum_{i=1}^{N} \sum_{j \neq i}^{N} \mathbf{F}_{ij} = \mathbf{0} \in \mathbb{R}^3,$$

implying the conservation of linear momentum, a classical result that neglects three-body forces. In Chapter 3 we utilised this conserved quantity to reduce the degrees of freedom, which manifested itself as translational invariance in the configuration.

Given forces governed by pairwise interactions that depend on the displacements of positions, namely $U_2 : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ with:

$$U_2 = \tilde{U}_2(\mathbf{q}_i - \mathbf{q}_j) \in \mathbb{R},$$

this will give rise to linear momentum conservation, but not necessarily angular mo-

mentum conservation. So it is usual to have forces that depend only on the interatomic distance $r_{ij} = |\mathbf{q}_j - \mathbf{q}_i| \in \mathbb{R}$. This is where the term central potential is used to describe potentials of the form $\Psi_2 : \mathbb{R} \to \mathbb{R}$ with:

$$U_2 = \Psi_2(r_{ij}) \in \mathbb{R}.$$

In this way, we consequently have Newton's third law in the weak form, in addition to the strong form holding which states the pairwise force $\mathbf{F}_{ij}$ is parallel to $\mathbf{q}_i - \mathbf{q}_j$. This is reiterated in equations (3.16) and (3.17).

This gives rise to an important question: what symmetries does a general interatomic potential U need to admit in order for us to conclude it is a central potential?

4.2 Theory

The configuration of a system of N atoms at positions $\mathbf{q}_i$, $i = 1, 2, \ldots, N$, is defined as a $3N$-dimensional vector $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N) \in \mathbb{R}^{3N}$. We note that these can provisionally be thought of as vertices of an N-gon, or an N-polyhedron, assuming that $\mathbf{q}_i \neq \mathbf{q}_j$ for $i \neq j$. The lengths of edges are distances between atoms, which we denote by

$$r_{ij} = |\mathbf{q}_j - \mathbf{q}_i|, \quad \text{for} \quad i, j = 1, 2, \ldots, N. \tag{4.2}$$

In this chapter, we study potential functions $U : \mathbb{R}^{3N} \to \mathbb{R}$ called central potential functions which satisfy certain symmetries as specified in Definition 1. These symmetries are: (i) translational invariance; (ii) rotational invariance; (iii) reflectional invariance; and (iv) parity for i,j identical atoms. An example of potential satisfying the assumptions in Definition 1 is

$$U(\mathbf{q}) = \sum_{i<j}^{N} \Psi_2(r_{ij}) + \sum_{i<j<k}^{N} \Psi_3(r_{ij}, r_{ik}, r_{jk}), \tag{4.3}$$

where $\Psi_2 : [0,\infty) \to \mathbb{R}$ and $\Psi_3 : [0,\infty)^3 \to \mathbb{R}$ are two-body and three-body potentials which depend on distances between atoms.

Definition 1. A function $U : \mathbb{R}^{3N} \to \mathbb{R} \cup \{\pm\infty\}$ is called a *central potential function* provided that it takes finite values on the subset

$$\Omega = \left\{ \mathbf{q} \in \mathbb{R}^{3N} \;\middle|\; \mathbf{q} = \{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_N\} \text{ with } \mathbf{q}_i \neq \mathbf{q}_j \text{ for } i \neq j \right\}$$

and for any $\mathbf{q} \in \Omega$, it satisfies:

(i) $U(\mathbf{q}_1 + \mathbf{c}, \mathbf{q}_2 + \mathbf{c}, \ldots, \mathbf{q}_N + \mathbf{c}) = U(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N)$ for all translations $\mathbf{c} \in \mathbb{R}^3$,

(ii) $U(R\,\mathbf{q}_1, R\,\mathbf{q}_2, \ldots, R\,\mathbf{q}_N) = U(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N)$ for all rotations $R \in \mathrm{SO}(3)$,

(iii) $U(Q\,\mathbf{q}_1, Q\,\mathbf{q}_2, \ldots, Q\,\mathbf{q}_N) = U(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N)$ for all reflections Q satisfying that all points $\mathbf{q}_i$, $i = 1, 2, \ldots, N$, lie on one side of the plane of reflection,

(iv) $U(\mathbf{q}_1, \ldots, \mathbf{q}_i, \ldots, \mathbf{q}_j, \ldots, \mathbf{q}_N) = U(\mathbf{q}_1, \ldots, \mathbf{q}_j, \ldots, \mathbf{q}_i, \ldots, \mathbf{q}_N)$ for any $i, j = 1, 2, \ldots, N$.

The symmetries considered in Definition 1 are satisfied by other generalizations of the example potential (4.3), which include n-body terms depending only on the distances (4.2) between atoms. In fact, the symmetries (i)-(iv) imply that the potential $U : \mathbb{R}^{3N} \to \mathbb{R}$ can be written as a function of distances. We have the following theorem which we prove in Section 4.3.

Theorem 1. *A central potential function $U : \mathbb{R}^{3N} \to \mathbb{R}$ can be written as*

$$\phi : [0,\infty)^{N(N-1)/2} \to \mathbb{R},$$

where the $N(N-1)/2$ inputs are interpreted as the set of all pairwise distances (4.2) *between atoms.*

Considering $N = 2$, Theorem 1 states that a central potential function U of 6 variables can be written as a function ϕ of 1 variable, r_{12}. Consequently, Theorem 1 reduces

the dimensionality of the potential U for any $N < 7$. If $N = 7$, then we have $3N = N(N-1)/2 = 21$ and the 21-dimensional state space $\mathbb{R}^{3N}$ corresponds to the 21 distance variables (4.2). Since the dimension of the state space scales as $O(N)$ and the number of distances scales as $O(N^2)$, Theorem 1 can be further improved by considering only a subset of the distance variables (4.2). In Section 4.3, we also prove the following result.

Theorem 2. *Let $N \geq 4$. Then a central potential function $U : \mathbb{R}^{3N} \to \mathbb{R}$ can be written as*

$$\phi : [0, \infty)^{4N-10} \to \mathbb{R},$$

where the $(4N-10)$ inputs are a subset of the set of all pairwise distances (4.2).

Considering $N = 4$ and $N = 5$, we have $4N - 10 = 6$ and $4N - 10 = 10$, respectively. In particular, Theorems 1 and 2 state the same conclusion for $N = 4$ and $N = 5$. Theorem 2 improves the result of Theorem 1 as $4N - 10 < N(N-1)/2$ for $N > 5$. Theorem 2 is a dimensionality reduction to the $3N$ positions for $N < 10$. We will prove Theorems 1 and 2 together in Section 4.3 by considering the cases $N = 2$, $N = 3$, $N = 4$, $N = 5$ and $N > 5$.

Applying Theorem 2 to our example potential (4.3), we observe that it reduces the number of independent variables for $N > 5$. In particular, while function ϕ constructed in the proof of Theorem 2 depends only on distances (4.2), it is not given in the form (4.3).

In addition to central potential functions satisfying conditions in Definition 1, there are potentials to which Theorems 1 and 2 are not applicable. For example, if the potential U corresponds to an external non-uniform field U_1, then we have

$$U(\mathbf{q}) = \sum_{i=1}^{N} U_1(\mathbf{q}_i)$$

and U will neither satisfy the conditions in Definition 1, nor will it be possible to

write as a function of pairwise distances (4.2). Assuming that there is no external field present and that we have a system of N identical atoms interacting (i.e. U satisfies condition (iv) in Definition 1), then we can formally write it as a sum of the n-body interactions for $2 \leq n \leq N$ in the form

$$U(\mathbf{q}) = \sum_{i<j}^{N} U_2(\mathbf{q}_i, \mathbf{q}_j) + \sum_{i<j<k}^{N} U_3(\mathbf{q}_i, \mathbf{q}_j, \mathbf{q}_k) + \dots \\ \dots + U_N(\mathbf{q}_1, \dots, \mathbf{q}_N), \quad (4.4)$$

where we can naturally think about n-polyhedrons of atoms as the input to the potential function, but these are fixed in space and a natural assumption is that given this input, it should not matter where we fix this polyhedron (leading to translational invariance (i)), or how we orient this polyhedron (rotational invariance (ii)). One slightly more subtle assumption, is that we should be allowed to reflect our polyhedron in any plane that keeps the polyhedron on one side (reflectional symmetry (iii)).

It is worth noting that although these symmetries are naturally understood, they are powerful from a Hamiltonian point of view. Noether's theorem says that each continuous symmetry gives rise to a corresponding conserved quantity (in a closed system), for example, translational invariance gives rise to conserved linear momentum (which is a consequence of reciprocity of forces). Therefore, we can intuitively understand that functions obeying these symmetries should only rely on distances. In the next section, we provide a proof of this conclusion, where we also show that a proper subset of pairwise distances for $N > 5$ can be used to describe the potential function U.

4.3 Proofs of Theorems 1 and 2

We prove Theorems 1 and 2, central potential theorems (CPTs), together by considering the cases $N = 2$, $N = 3$, $N = 4$ and $N = 5$, followed by an inductive argument for $N > 5$. We define displacement vectors by

$$\boldsymbol{\Delta}_{ij} = \mathbf{q}_j - \mathbf{q}_i, \quad \text{for} \quad i, j = 1, 2, \ldots, N, \tag{4.5}$$

i.e. we have $r_{ij} = |\boldsymbol{\Delta}_{ij}|$.

Let us start with the case $N = 2$. We define function $\phi : [0, \infty) \to \mathbb{R}$ by

$$\phi(s) = U(\mathbf{0}, s\hat{\mathbf{k}}) = U(0, 0, 0, 0, 0, s), \tag{4.6}$$

where $\hat{\mathbf{k}}$ is a unit vector in the direction of the positive z axis and $\mathbf{0} = [0, 0, 0]$. Given atom positions $\mathbf{q}_1, \mathbf{q}_2 \in \mathbb{R}^3$, we translate the configuration to position atom 1 at the origin. Using symmetry (i) in Definition 1, we have $U(\mathbf{q}_1, \mathbf{q}_2) = U(\mathbf{0}, \boldsymbol{\Delta}_{12})$. We then rotate the axes using rotation $R_1 \in \mathrm{SO}(3)$ such that the displacement vector connecting the two atoms is aligned with the positive z axis, giving $R_1\boldsymbol{\Delta}_{12} = r_{12}\hat{\mathbf{k}}$, while maintaining $R_1\mathbf{0} = \mathbf{0}$. Using symmetry (ii) in Definition 1, we have

$$U(\mathbf{q}_1, \mathbf{q}_2) = U(\mathbf{0}, \boldsymbol{\Delta}_{12}) = U(\mathbf{0}, r_{12}\hat{\mathbf{k}}) = \phi(r_{12}),$$

where the last equality follows from our definition (4.6). This concludes the proof of Theorem 1 for $N = 2$.

4.3.1 The case $N = 3$

Given atom positions $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3 \in \mathbb{R}^3$, we consider the function $U(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$. Using symmetry (i) in Definition 1, we translate the configuration to position atom 1 at the

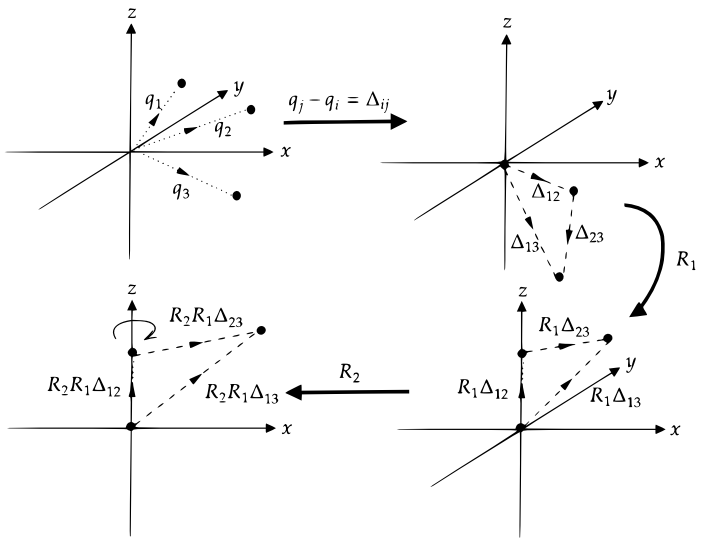


Figure 4.1: *A schematic of the constructive method in aid of the proof for the case* $N = 3$.

origin and consequently, we have

$$U(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) = U(\mathbf{0}, \mathbf{\Delta}_{12}, \mathbf{\Delta}_{13}).$$

Given that we have three axes to rotate around, we can always find a rotation R_1 such that $R_1\Delta_{12} = r_{12}\hat{\mathbf{k}}$, as we did in the $N = 2$ case. Using symmetry (ii), we have

$$U(\mathbf{0}, \mathbf{\Delta}_{12}, \mathbf{\Delta}_{13}) = U(\mathbf{0}, r_{12}\hat{\mathbf{k}}, R_1\mathbf{\Delta}_{13}).$$

We note that we can find a rotation R_2 about the z axis which rotates the triangle defined by the transformed atom positions $\mathbf{0}$, $r_{12}\hat{\mathbf{k}}$ and $R_1\mathbf{\Delta}_{13}$ to a planar triangle in the x-z plane, as demonstrated in Figure 4.1. Using symmetry (ii) again, we have

$$U(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) = U(\mathbf{0}, r_{12}\hat{\mathbf{k}}, R_2R_1\mathbf{\Delta}_{13}).$$

However the key point is that $R_2R_1\mathbf{\Delta}_{13}$ is uniquely defined by the triangle with lengths r_{12}, r_{13} and r_{23}, the angles of which can be calculated using the cosine rule,

i.e. $R_2R_1\boldsymbol{\Delta}_{13}$ can be expressed as

$$\left[\sqrt{r_{13}^2-\left(\frac{r_{13}^2+r_{12}^2-r_{23}^2}{2r_{12}}\right)^2}\,,\,0\,,\,\frac{r_{13}^2+r_{12}^2-r_{23}^2}{2r_{12}}\right]. \tag{4.7}$$

Therefore there exists function $\phi:[0,\infty)^3\to\mathbb{R}$ such that $U(\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3)=\phi(r_{12},r_{13},r_{23})$, for any $\mathbf{q}_1$, $\mathbf{q}_2$ and $\mathbf{q}_3$, confirming Theorem 1 for $N=3$.

4.3.2 The case $N=4$

Given atom positions $\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3,\mathbf{q}_4\in\mathbb{R}^3$, these can be thought of defining the vertices of a tetrahedron (or if co-planar a quadrilateral). Following similar steps as in the case $N=3$ in Section 4.3.1, we translate atom 1 to the origin, apply rotation R_1 to orient displacement vector $\boldsymbol{\Delta}_{12}$ with the positive z axis, then do a second rotation R_2 that fixes the triangle formed by the vertices of atoms 1, 2 and 3 in the x-z plane. As in Section 4.3.1, we have

$$U(\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3,\mathbf{q}_4)=U(\mathbf{0},r_{12}\hat{\mathbf{k}},R_2R_1\boldsymbol{\Delta}_{13},R_2R_1\boldsymbol{\Delta}_{14}).$$

Using equation (4.7), we know that $R_2R_1\boldsymbol{\Delta}_{13}$ is determined entirely by distances r_{12}, r_{13} and r_{23}. All that remains to be shown is that $R_2R_1\boldsymbol{\Delta}_{14}$ is determined by pairwise distances. We note that the triangle formed by atoms 1, 2 and 3 (denoted as ABC in the lower part of our illustration of the proof in Figure 4.2) is uniquely determined (after orienting one side with the positive z axis). Consequently, this fixes the side BC. On the other hand the triangle BCD is uniquely determined (as one side BC is fixed) by distances r_{23}, r_{24} and r_{34}. These can be thought of as two triangles which can rotate around a hinge BC, so to determine the vector $R_2R_1\boldsymbol{\Delta}_{14}$, we necessarily need the final distance r_{14} that gives the angle between the planes containing triangles ABC and BCD (two configurations are illustrated in Figure 4.2). If triangles ABC and

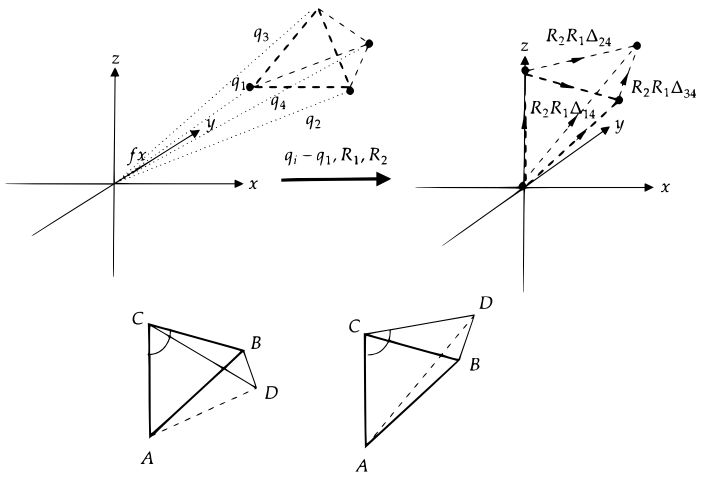


Figure 4.2: *A schematic of the constructive method in aid of the proof for $N = 4$. For clarity we have only highlighted the additional three displacement vectors, though the triangle formed by vertices $\{1, 2, 3\}$ lying in the $x - z$ plane is the same as in Figure* 4.1*.*

BCD are co-planar, the set of all pairwise distances, with this orientation, will give a unique description of $R_2R_1\mathbf{\Delta}_{14}$. If these triangles are not co-planar, this final distance gives two possible vectors for $R_2R_1\mathbf{\Delta}_{14}$. These correspond to a unique $R_2R_1\mathbf{\Delta}_{14}$ and the copy obtained by reflection in the plane containing triangle ABC. However by property (iii) we know that if we reflect in the plane containing ABC with a matrix denoted Q, then

$$\begin{aligned} &U(\mathbf{0}, r_{12}\hat{\mathbf{k}}, R_2R_1\mathbf{\Delta}_{13}, R_2R_1\mathbf{\Delta}_{14}) \\ &\quad = U(\mathbf{0}, r_{12}\hat{\mathbf{k}}, R_2R_1\mathbf{\Delta}_{13}, QR_2R_1\mathbf{\Delta}_{14}) \end{aligned}$$

Therefore there exists function $\phi : [0, \infty)^6 \to \mathbb{R}$ such that
$U(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4) = \phi(r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34})$, for any $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ and $\mathbf{q}_4$, confirming Theorem 1 for $N = 4$.

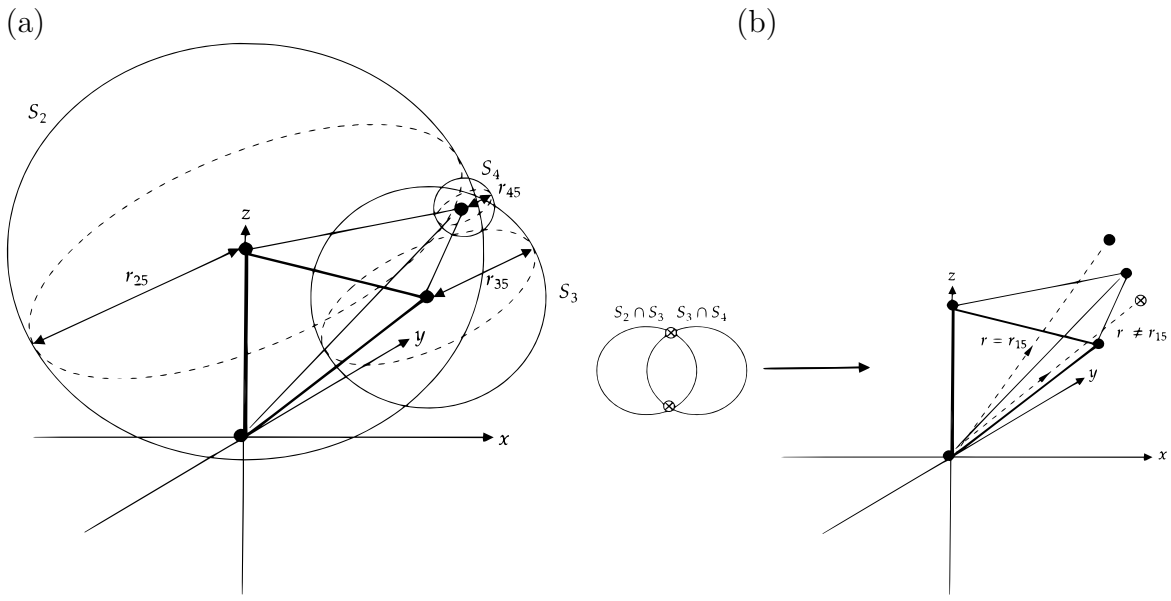


Figure 4.3: (a) *The intersection of* 3 *spheres, based on three known centres and radii are used to position an additional vertex.*
(b) *Here a fourth vertex, chosen non co-planar to the three vertices used to construct the spheres previously, is used to uniquely determine the fourth vertex position.*

4.3.3 The case $N = 5$

To proceed in this case, we note that any N vertex polyhedron can be made by adding a single vertex to an $N-1$ polyhedron or polygon (in the case where all other points are co-planar). The task at hand, as in the case $N = 4$ in Section 4.3.2, is being able to determine the displacement vectors once we have translated and rotated the configuration such that $R_2 R_1 \boldsymbol{\Delta}_{12} = r_{12}\hat{\mathbf{k}}$ is aligned with the positive z axis.

An $N = 5$ polyhedron can be constructed from either adding a vertex onto a pre-existing $N = 4$ polyhedron (at most 3 points are co-planar), or an $N = 4$ polygon (where all points are co-planar). In the first case, we may take any 3 vertices on the pre-existing polyhedron: call these vertices the transformed positions of atoms $2, 3, 4$ (by property (iv) in Definition 1). If we know r_{25}, then this fifth vertex must lie on a sphere of radius r_{25}, with the transformed position of atom 2 as the centre: we denote this S_2. Similarly, we construct S_3 and S_4 as spheres of radii r_{35} and r_{45} respectively. This is illustrated in Figure 4.3(a). The fifth vertex lies at the intersection of three

spheres S_2, S_3 and S_4, which contains at most two points. If it contains exactly two points, then we need another distance r_{15} (which is the distance from the vertex in the pre-existing polyhedron that was not used as a centre of spheres S_2, S_3 or S_4) to determine which of those two positions is correct, see Figure 4.3(b). In this way: 4 more distances are used to specify all of the vertices of the $N = 5$ polyhedron. Therefore, there exists function $\phi : [0, \infty)^{10} \to \mathbb{R}$ such that $U(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4, \mathbf{q}_5) = \phi(r_{12}, r_{13}, r_{14}, r_{15}, r_{23}, r_{24}, r_{25}, r_{34}, r_{35}, r_{45})$, for any $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$, $\mathbf{q}_4$ and $\mathbf{q}_5$.

To arrive at this conclusion, we used an assumption that no four points are co-planar. If this is not the case, then we need less distances for the specific configuration. For example, if the pre-existing 4 vertices are co-planar: utilising the sphere approach for any three of those vertices will result again in two possible positions for vertex 5, however using the pairwise distance between this vertex and the new vertex gives no information, as the fourth point lies on the plane of symmetry formed by the spheres. In this case we use property (iii), considering the reflective symmetry about this plane to argue that we have determined all displacement vectors with this orientation uniquely up to a reflection in the plane containing vertices 1, 2, 3 and 4. In this case, we do not need the fourth distance mentioned above, and evaluating $U(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4, \mathbf{q}_5)$ is possible with the 9 pairwise distances. The tenth distance is also not needed if the intersection of spheres S_2, S_3 and S_4 is exactly equal to one point (when vertices 2, 3 and 4 are co-linear). Thus we have proven Theorems 1 and 2 in the case $N = 5$.

4.3.4 The case $N > 5$

We inductively prove that a similar setup as in the $N = 5$ case in Section 4.3.3 works by constructing polyhedra of higher order by the addition of a new vertex. Say that the $N-1$ case required the set of distances D_{N-1} to evaluate $U(\mathbf{q}_1, \ldots, \mathbf{q}_{N-1})$, where $|D_{N-1}| = 4(N-1) - 10$. Which is true for the base case of $N = 5$.

The most general case to consider is when we have an $N-1$ polyhedron before we introduce the new vertex. In this case, any three of the $N-1$ vertices can be chosen, say i, j, k. The three distances r_{iN}, r_{jN}, r_{kN} are used to create three intersecting spheres and two potential positions for vertex N. We use vertex l, which is not co-planar to i, j, k and distance r_{lN} determines this position uniquely. Therefore the required set of distances to evaluate $U(\mathbf{q}_1, \ldots, \mathbf{q}_N)$ is $D_N = D_{N-1} \cup \{r_{iN}, r_{jN}, r_{kN}, r_{lN}\}$, i.e. we need $4N-10$ pairwise distances. If there are at least four co-planar points, then we only need 3 additional pairwise distances (so we would only need $4N-11$ pairwise distances). Since the inductive step holds for all N, and it works for the base case of $N = 5$, this concludes our proofs of Theorems 1 and 2 for all N.

4.4 Corollaries

Corollary 1 (Minimal distance set)**.** *Given a central potential function* $U : \mathbb{R}^{3N} \to \mathbb{R}$*, where* $\mathbf{q} = \{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_N\}$*, with* $\mathbf{q}_i \neq \mathbf{q}_j \ \forall i \neq j$*, then* U *can be written as* $\phi : (\mathbb{R}^+)^k \to \mathbb{R}$ *with* $k = 6+3(N-4)$ *for* $N \geq 4$*, given a correct initial orientation that can always be found.*

Proof. We first note that given N vertices, one can always find a triple of vertices $\{i, j, k\}$ that form a triangular face (or sub-face) of the polyhedron, that has the property that it lies on the surface of the convex hull formed from the N points. The key point is that all other points will lie on one side of this triangular face by the given convexity. An example is having 5 points, but the convex hull formed is a tetrahedron with one interior vertex enclosed: in this case 4 triples are admissible by this criterion, and each form one of the outer triangular faces.

Every set of vertices has at least 2 triples satisfying this property: a minimal example is given by a 7 vertex polyhedron which is formed by two tetrahedrons sharing one vertex - only the 2 outer triangular bases form the required triple.

We propose a practical way, to find such a triple, which can then be used to

generate the 3d convex hull in $O(Nh_N)$ time, similar to planar Jarvis marching, where N is the number of points, and h_N is the number of points on the convex hull. Given N random points in $\mathbb{R}^3$, one can perform a max search to locate the point with the largest z coordinate: which takes $N-1$ flops. Taking the plane $z = c_1$, where c_1 is the maximum z component of all the points: corresponding to the point (a_1, b_1, c_1) - this is vertex labelled i.

From here, we can pivot the plane on this point in any direction, for simplicity pick the direction such that the unit normal is in the $x-z$ plane. We locate the next vertex by minimising the $x-z$ gradient between vertex i, and all other $N-1$ points, by calculating $|\Delta z/\Delta x|$. The point that forms the shallowest gradient (the sign is irrelevant hence we take the modulus) with vertex i must lie on the convex hull. Intuitively, if we were to drop an infinite rod from a horizontal orientation and it met vertex i, it would have to pivot about this point in the $x-z$ plane before it met a point of first contact. This is denoted vertex j, and this procedure takes a further $4(N-1)$ flops to calculate all of the gradients and taking the modulus - then we perform a minimum search for an additional $N-2$ flops: totalling $5N-6$ flops.

Finally, we form the equation of the plane that contains vertex i and j with the assumption that the unit normal lies in the $x-z$ plane: $n = (n_1, 0, n_3)$. Using the fact that 2 points lie on the plane, we simply find n, which requires 11 flops. Thus we can define the plane Π, and the infinite line L that connects vertex i and j. As per figure 4.4, from all remaining $N-2$ points calculate the shortest distance to L and Π, we denote these d_1 and d_2 respectively. These distances, from point $m = 1, \ldots, N-2$, are given by standardised formulas:

$$d_1 = \sqrt{\frac{|\mathbf{q}_i - \mathbf{q}_m|^2 |\mathbf{q}_i - \mathbf{q}_j|^2 - [(\mathbf{q}_i - \mathbf{q}_m) \cdot (\mathbf{q}_j - \mathbf{q}_i)]^2}{|\mathbf{q}_i - \mathbf{q}_j|^2}} \tag{4.8}$$

$$d_2 = |n_1 q_k^x + n_3 q_k^z + D|, \tag{4.9}$$

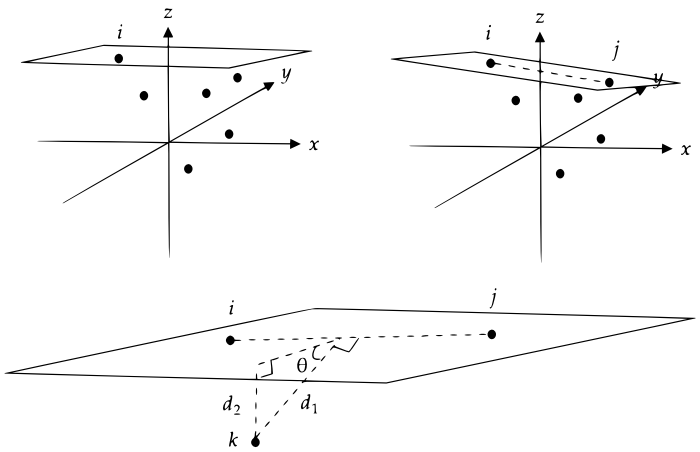


Figure 4.4: *This figure illustrates the proposed procedure to construct the first triangular face of a convex hull of points in 3d by locating vertices i, j and k.*

where $D = n_1 q_i^x + n_3 q_i^z = n_1 q_j^x + n_3 q_j^z$ by construction and is calculated in defining $\Pi : n_1 x + n_3 z - D = 0$.

We note that equation (4.8) requires an initial calculation of $|\mathbf{q}_i - \mathbf{q}_j|^2$ which requires 8 flops. For each point $m = 1, \ldots, N-2$, we require an additional 18 flops to calculate d_1. Similarly, use of equation (4.9) requires 5 flops for each of the $N-2$ values of d_2. For this stage of the process we necessarily have $8 + 23(N-2)$ flops. The key point here is that we wish to calculate the minimum angle θ, illustrated in figure 4.4, as this means we have to turn the plane Π until we hit the first vertex, denoted k. This forms the final triple $\{i, j, k\}$, which by construction has all other $N-3$ points to one side of it. We lastly note that $d_2/d_1 = \sin(\theta)$, so we require the minimum of the ratio d_2/d_1, which takes a further $N-2$ flops, in addition to the minimum search of $N-3$ flops. Enumerating this, the final vertex requires an additional $25N-32$ flops (for $N > 3$). In total the procedure for finding the first face is $31N - 39$ flops, which is all we need for this work. Though one can procedurally carry this out using one of the other vertices: j or k to start the process again, doing this until the entire convex hull is reconstructed in $O(Nh_N)$ time.

The procedure above is fairly crude but gives a practical way to reconstruct this first triple $\{i, j, k\}$. Given this triple $\{i, j, k\}$, we note by property (iv) we can label

these are vertices $\{1, 2, 3\}$. When following the procedure in the proof of the previous theorem: we can orient this triangular face such that it lies entirely in the $x-z$ plane, and by construction, all of the other vertices are either co-planar in the $y = 0$ plane, or lie entirely in $y > 0$ or $y < 0$.

When placing vertex 4, we can use the distances $\{r_{14}, r_{24}, r_{34}\}$ to construct the three intersecting spheres, as argued before in the $N = 4$ case. The choice of intersection is symmetric in the plane $y = 0$ so we can pick any such intersection point: this will either have $y = 0$, $y > 0$ or $y < 0$. Given this, we can use distances $\{r_{1p}, r_{2p}, r_{3p}\}$ to place the p^{th} vertex (for $p > 4$) noting that the side has already been determined by the property that all vertices should lie to one side of $y = 0$; namely $y \leq 0$ or $y \geq 0$ for all vertices. In this case, we don't need a fourth distance to uniquely position any of the vertices by exploiting the symmetry of property (iii).

To conclude, given one of these triangular faces on the convex hull - we can tether every subsequent vertex with the additional 3 pairwise distances formed from this triangle to the point, forming the vector of distances:
$\mathbf{r} = (r_{12}, r_{13}, r_{14}, \ldots, r_{1N}, r_{23}, r_{24}, \ldots, r_{2N}, r_{34}, r_{35}, \ldots, r_{3N})$. This means that potential U can be written as $\phi(\mathbf{r}) : (\mathbb{R}^+)^{6+3(N-4)} \to \mathbb{R}$ for $N \geq 4$. ■

Remark. *The result would come to no surprise for anyone familiar with potentials, it makes intuitive sense that $N(N-1)/2$ pairwise distances provide an adequate framework to work in. Given the $3N$ general positional degrees of freedom, translational invariance reduces three of these, and rotational invariance should also reduce this further. It is however not immediately obvious that given these symmetries, $3N - 6$ (for $N \geq 4$) independent pairwise distances characterise this, which is what we have shown: we will explicitly construct the function ϕ later on in Section 4.6.*

Corollary 2 (Decomposition of many-body potentials)**.** *Given a system of N identical interacting atoms, governed by a potential function $\phi : (R^+)^{N(N-1)/2} \to \mathbb{R}$ in which:*

(i) ϕ depends only on pairwise distances r_{ij} where $i, j = 1, ..., N$ with $i \neq j$,

(ii) ϕ depends only on $n \leq N$-body interactions,

then

$$\phi(r_{12}, r_{13}, \ldots, r_{N-1\,N}) = \begin{cases} \sum_{i<j}^{N} \Psi_2(r_{ij}) & \text{if } n = 2 \\ \sum_{i=1}^{\binom{N}{n}} \Psi_n(\mathbf{r}_i) & \text{if } n \geq 3 \end{cases}$$

where Ψ_n is an arbitrary function, and $\mathbf{r}_i$ is a vector of, 3 distances for $n = 3$ and $4n - 10$ distances for $n \geq 4$, which define each distinct (up to rotation/reflection) n polygon/polyhedron that can be formed from a set of N vertices of a polyhedron.

Proof. We have seen and used the fact that N identical atoms can form the vertices of either an N polyhedron, or N polygon. The potential function only depends on pairwise distances r_{ij}, of which there are $N(N-1)/2$. If we think about a planar embedding of this polyhedron (which can be done for convex polyhedron by Tutte's Embedding Theorem), we can connect all of these embedded vertices to form a fully connected graph. The line connecting any two vertices represents the embedded distance between the vertices.

In the case of N atoms and a potential that depends solely on $n = 2$ fixed body interactions (a pair potential), then fixing each line is equivalent to fixing the entire polyhedron. A single pairwise potential energy contribution is calculated with the distance between two atoms, and does not take into consideration the positions of other atoms, which leads to the natural conclusion that $\phi(r_{12}, r_{13}, \ldots, r_{N-1\,N}) = \sum_{i<j}^{N} \Psi_2(r_{ij})$.

One can then think about $n = 3$-body interactions, which necessarily depend on each fixed triangle, formed by any three of the N vertices. Fixing each triangle is equivalent to fixing the entire polyhedron, as shown in figure 4.5 (with a plane embedding onto an irregular pentagon). Given each potential energy contribution requires a fixed triangle (as the potential is only distance dependent by property (i)); we necessarily have that our potential that depends only on three-body interactions looks

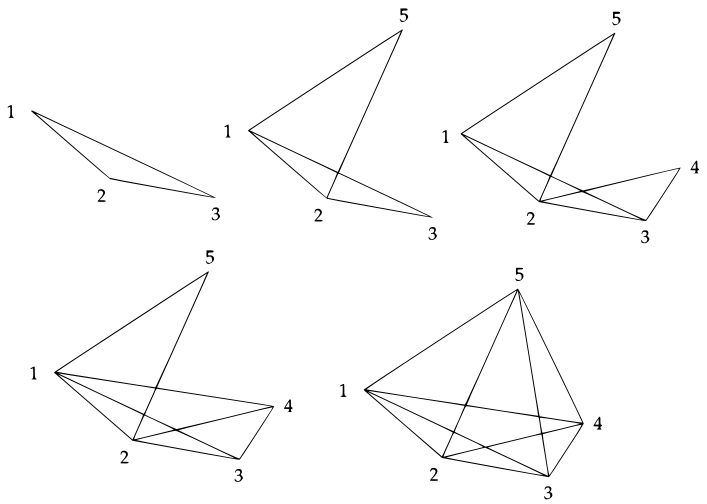


Figure 4.5: *Here we have embedded the $N = 5$ polyhedron onto a planar graph, and in fixing each triangle, we can reconstruct the full configuration. We note that there are $\binom{N}{3}$ triangles that need to be fixed, with only 5 shown here as the others overlap with the 2d representation. The other triangles are used to fix the 3d distances between vertices.*

like $\phi(r_{12}, r_{13}, \ldots, r_{N-1\,N}) = \sum_{i=1}^{\binom{N}{3}} \Psi_3(\mathbf{r}_i)$, where $\mathbf{r_i} \in (\mathbb{R}^+)^3$ is a vector that contains a triple $\{r_{ij}, r_{jk}, r_{ki}\}$ of distances in which the index pairs (which we can transpose as $r_{ij} = r_{ji}$) form a 3 cycle (ijk).

In a similar vein we can think about general n-body interactions, noting that fixing each n polygon/polyhedron, fixes the entire configuration. By a similar reasoning we have that $\phi(r_{12}, r_{13}, \ldots, r_{N-1\,N}) = \sum_{i=1}^{\binom{N}{n}} \Psi_n(\mathbf{r}_i)$, where we will work on defining $\mathbf{r}_i$, as again - an arbitrary collection of distances does not define a connected polyhedron. We start by picking a distance r_{ij} for some i, j, this fixes atoms i and j relative to each other (by property (i) the orientation does not matter). We wish to form any triangle connecting three vertices, in this case i, j, k, so from before, we must pick r_{jk} and r_{ki}. This uniquely defines a triangle up to a reflection in the side connecting atoms i and j as two circles C_{ik} and C_{jk} of radii r_{jk} and r_{ik}, with centres: vertex corresponding to atom i and vertex corresponding to atom j respectively, can be created. The intersection of which gives at most two points (one if the third vertex is co-linear, no intersections is precluded here as the third vertex does exist), and each

choice produces an equivalent triangle, consequently, a potential energy contribution. We say that $d_3(i,j,k) = \{r_{ij}, r_{jk}, r_{ki}\}$.

To specify the fourth point on the polyhedron, we need an additional three distances r_{il}, r_{jl}, r_{kl} for a similar reasoning to the previous theorem, the intersection of three spheres with distinct centres form two points, each of which gives the same potential energy. Note we have to pick the distances from atoms i,j,k to this new atom l. In this way, to define all four point polyhedrons we need sets which look like $d_4(i,j,k,l) = d_3(i,j,k) \cup \{r_{il}, r_{jl}, r_{kl}\}$ - these form the components of each $\mathbf{r}_i$. And the potential function should be a sum over $\binom{N}{4}$ such sets as we can choose vertices i,j,k,l. In the case where 2 of the original i,j,k in addition to l vertices are co-linear, only two distances are needed as the circles intersect at 1 point.

Similar to before, there are more redundancies past this owing to the linear dependence of distances, not including the cases where three atoms are co-linear. One requires, generally, 4 more distances to pin point the additional vertex; so when forming $d_6(i,j,k,l,m,n)$ there are multiple choices of sets of distances which will always work, namely

$$\begin{aligned} d_6(i,j,k,l,m,n) &= d_5(i,j,k,l,m) \cup \{r_{in}, r_{jn}, r_{kn}, r_{ln}\} \\ &= d_5(i,j,k,l,m) \cup \{r_{in}, r_{jn}, r_{kn}, r_{mn}\} \\ &= d_5(i,j,k,l,m) \cup \{r_{in}, r_{jn}, r_{ln}, r_{mn}\} \\ &= d_5(i,j,k,l,m) \cup \{r_{in}, r_{kn}, r_{ln}, r_{mn}\} \\ &= d_5(i,j,k,l,m) \cup \{r_{jn}, r_{kn}, r_{ln}, r_{mn}\}, \end{aligned}$$

as we have pinned down the previous vertices, we can choose any four to tether the fifth vertex. Hence for n-body interactions, the size of $\mathbf{r}_i$ is $4n-10$ for $n \geq 4$ in the worst case scenario there are no three points that are co-linear during any of the

construction. ■

Corollary 3. *The many-body potential given in equation* (4.4) *that satisfies properties (i)–(iv) can be written as*

$$U(\mathbf{q}) = \sum_{i<j}^{N} \Psi_2(r_{ij}) + \sum_{n=3}^{N} \sum_{i=1}^{\binom{N}{n}} \Psi_n(\mathbf{r}_i),$$

where $\mathbf{r}_i$ *is a vector containing the appropriate set of* $4n-10$ *distances used to define each* $\binom{N}{n}$ *n-polyhedron.*

Proof. Given U satisfies the symmetric properties (i)–(iv), Theorem 1 allows us to conclude that U must be parameterisable by pairwise distances r_{ij}. We then use Corollary 2 to decompose each n-body term in equation (4.4) into sums over n-polyhedra, parameterising each contribution with the $4n-10$ distances as required. ■

4.5 Two dimensional planar case

To highlight the case where we are adding an additional vertex to a set of $N-1$ co-planar vertices, and the additional point is itself co-planar; we are essentially in the remit of a 2d version of the general theorems present before in 3d.

Again, we will require a single distance r_{ij} to fix atoms i and j in some orientation, and the triple of distances $\{r_{ij}, r_{jk}, r_{ik}\}$ to fix $N=3$ atoms. This is similar to the 3d case as three points are always co-planar. We can similarly rotate to place the entire configuration in the $x-z$ plane with $R_2R_1\mathbf{\Delta}_{12}$ aligning with the positive z axis.

In this case, at most three distances are required to place an additional vertex after the initial triangle is placed, due to the fact we have 3 intersecting circles (as opposed to spheres) with distinct centres which will intersect at 1 point. We can then

reason the equivalent form of Theorem 2 is that for $N \geq 3$:

$$\phi : (\mathbb{R}^+)^{3N-6} \to \mathbb{R}.$$

We can use the first two stages of our proposed convex hull algorithm, namely finding the point with the maximum z coordinate in $N-1$ flops, then finding the shallowest $x-z$ gradient in $5N-6$ flops for a total of $6N-7$ flops. This identifies two vertices we can label $\{1,2\}$ which form a line segment on the convex hull of 2d points. In this way, we can formulate the general 2d minimum distance description of ϕ. With the orientation along the positive z axis, given that this is a co-planar configuration with vertices $\{1,2\}$ on the convex hull, all points must have $x \leq 0$ or $x \geq 0$.

In this way, only two distances are necessarily required to specify the additional vertex after we have a degree of freedom on which side to place the third vertex (this will define which side all other vertices must lie, but by property (iii) we have reflectional symmetry in the z axis by construction).

The equivalent form of Corollary 1 in 2d is that we can write $\phi : (\mathbb{R}^+)^k \to \mathbb{R}$ with $k = 1 + 2(N-2)$ for $N \geq 2$, given a correct initial orientation that can always be found. In this way the subset of pairwise distances required is $\mathbf{r} = \{r_{12}, r_{13}, r_{14}, \ldots, r_{1N}, r_{23}, r_{24}, \ldots, r_{2N}\}$.

Finally, we note that the trivial case where all atoms are co-linear requires exactly $1 + 2(N-2)$ pairwise distances in 2d or 3d: where we can pick any $\{i,j\}$ as vertices $\{1,2\}$, and the reflective symmetry of our potential is redundant as any two circles (or indeed spheres), intersect at precisely one point.

4.6 Applications

We begin this section by detailing how to construct the function ϕ in corollary 1. Given we have labelled the vertices of any triangular face of the convex hull formed

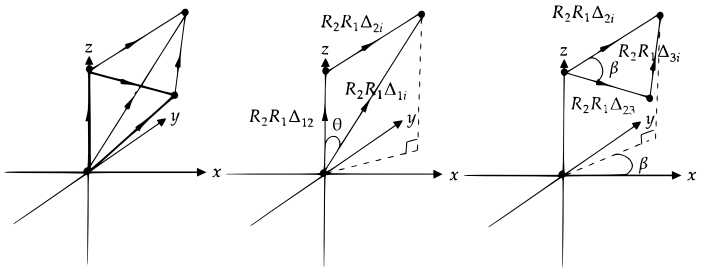


Figure 4.6: *In this figure we illustrate how one can form the vector* $\Delta_{1i}^{(2)}$ *with distances* $\{r_{12}, r_{23}, r_{2i}, r_{1i}, r_{3i}\}$ *by defining angles* θ *(indexed in this case as* θ_{21i}*) and* β *(indexed as* β_{32i}*). These are commonly referred to as polar and azimuthal angles in spherical polar systems. The starting configuration shown is taken after the translation* $\mathbf{q}_i - \mathbf{q}_1$*, rotation* R_1 *and rotation* R_2*.*

by the N vertices as $\{1, 2, 3\}$ and oriented this in the $x - z$ plane; forming the vector $R_2R_1\mathbf{\Delta}_{1i}$ takes only an additional three distances $\{r_{2i}, r_{1i}, r_{3i}\}$. Using figure 4.6 to illustrate, we can form the triangle with vertices $\{1, 2, i\}$, defining the θ_{21i} as the angle between vectors $R_2R_1\mathbf{\Delta}_{12}$ and $R_2R_1\mathbf{\Delta}_{1i}$: we have that

$$\cos(\theta_{21i}) = \frac{r_{2i}^2 - r_{12}^2 - r_{1i}^2}{-2r_{12}r_{1i}}. \tag{4.10}$$

We then deduce the z component of $R_2R_1\mathbf{\Delta}_{1i}$ is $r_{1i}\cos(\theta_{21i})$, whilst the distance of $r_{1i}\sin(\theta_{21i})$ is projected onto the $x - y$ plane.

Similarly the triangle formed by vertices $\{2, 3, i\}$ is used to define β_{32i} which is the angle between vectors $R_2R_1\mathbf{\Delta}_{23}$ and $R_2R_1\mathbf{\Delta}_{2i}$: we have that

$$\cos(\beta_{32i}) = \frac{r_{3i}^2 - r_{23}^2 - r_{2i}^2}{-2r_{23}r_{2i}}. \tag{4.11}$$

In this way we can explicitly write (using spherical coordinates)

$$R_2R_1\mathbf{\Delta}_{1i} = \begin{pmatrix} r_{1i}\sin(\theta_{21i})\cos(\beta_{32i}) \\ r_{1i}\sin(\theta_{21i})\sin(\beta_{32i}) \\ r_{1i}\cos(\theta_{21i}). \end{pmatrix}$$

We note that inversion of equations (4.10), (4.11) can give rise to two solutions in the range $[0, 2\pi)$, which may be a cause for concern. However, we have restricted all vertices to $y \leq 0$ and $y \geq 0$, which means we can restrict the range of azimuthal angle β to give one solution.

In this way, we have that

$$\begin{aligned} U(\mathbf{q}) &= U(\mathbf{0}, r_{12}\hat{\mathbf{k}}, R_2R_1\boldsymbol{\Delta}_{13}, \ldots, R_2R_1\boldsymbol{\Delta}_{1i}, \ldots, R_2R_1\boldsymbol{\Delta}_{1N}) \\ &= U(0, 0, 0, 0, 0, r_{12}, r_{13}\sin(\theta_{213}), 0, r_{13}\sin(\theta_{213}), \ldots, \\ &\qquad r_{1i}\sin(\theta_{21i})\cos(\beta_{32i}), r_{1i}\sin(\theta_{21i})\sin(\beta_{32i}), r_{1i}\cos(\theta_{21i}), \ldots, \\ &\qquad r_{1N}\sin(\theta_{21N})\cos(\beta_{32N}), r_{1N}\sin(\theta_{21N})\sin(\beta_{32N}), r_{1N}\cos(\theta_{21i})) \\ &= \phi(r_{12}, r_{13}, \ldots, r_{1N}, r_{23}, r_{24}, \ldots, r_{2N}, r_{34}, r_{35}, \ldots, r_{3N}), \end{aligned} \tag{4.12}$$

which explicitly defines function ϕ in terms of $6 + 3(N-4)$ distances (for $N \geq 4$) for a general potential function $U(\mathbf{q})$ satisfying properties (i)–(iv).

We turn our attention to force calculations under the assumption that these can be written as the gradient of positions: the force on atom i is given by

$$\mathbf{F}_i = -\nabla_i U(\mathbf{q}),$$

which for pair potentials ends up being a sum over all of the pairwise contributions.

With this in mind, it is notable that given a potential function that exhibits just translational invariance, such that $U_2(\mathbf{q}_i, \mathbf{q}_j) = \tilde{U}_2(\mathbf{q}_i - \mathbf{q}_j)$ for all pair contributions: we necessarily have that the force on atom i induced by atom j is:

$$\mathbf{F}_{ij} = -\nabla_i \tilde{U}_2(\mathbf{q}_i - \mathbf{q}_j),$$

and this functional form gives rise to reciprocity in the sense that action equals reaction (weak form of Newton's third law): $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$. Translational invariance

alone doesn't give rise to potentials adhering to the strong form of Newton's third law which dictates that pairwise forces act along the vector connecting both atoms, rotational invariance is required as well.

In fact, when we can write a potential in terms of pairwise distances $\phi(r_{ij})$: we necessarily have that

$$\mathbf{F}_{ij} = -\nabla_i \phi(r_{ij}) = -\frac{1}{r_{ij}} \frac{\partial \phi}{\partial r_{ij}}(r_{ij}) \mathbf{\Delta}_{ji},$$

where Newton's third law, both strong and weak form hold.

In the case of central potential $U(\mathbf{q})$, using equation (4.4) and Corollary 1 gives us that

$$\mathbf{F}_i = -\nabla_i U(\mathbf{q}) = -\nabla_i \phi(\mathbf{r}) = -\sum_{j \neq i}^{N} \frac{1}{r_{ij}} \frac{\partial \phi}{\partial r_{ij}}(\mathbf{r}) \mathbf{\Delta}_{ji}. \tag{4.13}$$

In an exact treatment: one has $N-1$ contributions to calculate in equation (4.13) for every $i = 1, \ldots, N$ atom, where typically the total force calculation forms the bottleneck of MD as an $O(N^2)$ process. Utilising the symmetry of $r_{ij} = r_{ji}$ and with no implementation of approximation schemes, means in this case one has to explicitly calculate $N(N-1)/2$ distinct contributions.

With $\mathbf{r} = (r_{12}, r_{13}, r_{14}, \ldots, r_{1N}, r_{23}, r_{24}, \ldots, r_{2N}, r_{34}, r_{35}, \ldots, r_{3N})$, many such derivatives are trivial now. For example, the term corresponding to $\partial\phi/\partial r_{mn} = 0$ where $m, n \notin \{1, 2, 3\}$.

In this way we can conclude that:

$$\mathbf{F}_i = \begin{cases} -\sum\limits_{j \neq i}^{N} \frac{1}{r_{ij}} \frac{\partial \phi}{\partial r_{ij}}(\mathbf{r}) \mathbf{\Delta}_{ji} & \text{if } i = 1, 2, 3 \\ -\sum\limits_{j=1}^{3} \frac{1}{r_{ij}} \frac{\partial \phi}{\partial r_{ij}}(\mathbf{r}) \mathbf{\Delta}_{ji} & \text{if } i \geq 4 \end{cases} \tag{4.14}$$

which leads to a total of $3N - 6$ distinct force contributions (for $N \geq 3$) owing to

the symmetry of $r_{ij} = r_{ji}$ and noting each contribution arises from $\partial\phi/\partial r_{mn}$. The process of calculating forces now scales linearly as $O(N)$, and directly comparing to the position dependent parameterisation of $N(N-1)/2$ contributions, we see viewing the problem in this way leads to equal or fewer contributions for $N \geq 3$. The price we pay for this linear scaling is that the partial derivatives become more of a handful, though this can be tackled fairly easily on a case by case basis. The advantage here is that once the derivatives are dealt with, you can do exact force calculations that scale linearly with the number of atoms, a result the requires no additional assumptions except that forces are governed by a central potential.

To demonstrate these theoretical results in the most practical way; we elucidate, with some generality, how one can implement this to known central potentials. For simplicity we will consider an N-atom pair potential though the methods of transforming the potential in terms of distances, and calculating forces, are general.

To start, typically in literature we are presented with a central pair potential of the form in equation (4.3), without the three-body term:

$$U(\mathbf{q}_1, \ldots, \mathbf{q}_N) = \sum_{i<j}^{N} \Psi_2(r_{ij}) = \sum_{i<j}^{N} \Psi_2(|\mathbf{q}_i - \mathbf{q}_j|).$$

Noting that we can use the minimal distance description given by equation (4.12) (after an appropriate relabelling such that atoms $\{1, 2, 3\}$ are vertices of a triangular face lying on the convex hull) we derive the set of substitutions (utilising the cosine rule to relates distances based on triangles illustrated in figure 4.6) for r_{ij} with $i \neq j = 1, \ldots, N$:

$$|\mathbf{q}_i - \mathbf{q}_j| = \sqrt{r_{1i}^2 + r_{1j}^2 - 2r_{1i}r_{1j}[A_{ij} + B_{ij} + C_{ij}]} \tag{4.15}$$

where

$$
\begin{aligned}
A_{ij} &= \sin(\theta_{21i})\sin(\theta_{21j})\cos(\beta_{32i})\cos(\beta_{32j}) \\
B_{ij} &= \sin(\theta_{21i})\sin(\theta_{21j})\sin(\beta_{32i})\sin(\beta_{32j}) \\
C_{ij} &= \cos(\theta_{21i})\cos(\theta_{21j}).
\end{aligned}
$$

It is worth noting if any $i, j = 1, 2$ then $A_{ij} = 0$ as angles $\theta_{21i} = \theta_{21j} = 0$. Similarly if $i, j = 1, 2, 3$ then $B_{ij} = 0$. In addition, as r_{ii}=0 for all i, we see expression (4.15) gives $|\mathbf{q}_1 - \mathbf{q}_i| = r_{1i}$, $|\mathbf{q}_2 - \mathbf{q}_i| = r_{2i}$ for all $i = 1, \ldots, N$.

Finally, from equations (4.10) and (4.11), we conclude that A_{ij} and B_{ij} depend on the set of distances $\{r_{12}, r_{1i}, r_{2i}, r_{1j}, r_{2j}, r_{23}, r_{3i}, r_{3j}\}$, whereas C_{ij} depends on a proper subset of these, namely $\{r_{12}, r_{1i}, r_{2i}, r_{1j}, r_{2j}\}$.

In a pedagogical case where $\Psi_2(r_{ij}) = 1/r_{ij}$, we could then make the substitutions given in equation (4.15) and calculate derivatives $\partial\Psi_2(r_{ij})/\partial r_{ij}$ in accordance with the force expression (4.14). If we take an arbitrary N we have

$$
\begin{aligned}
\phi &= \sum_{i=2}^{N} \frac{1}{r_{1i}} + \sum_{i=3}^{N} \frac{1}{r_{2i}} \\
&\quad + \sum_{i=3}^{N-1} \sum_{j=i+1}^{N} \frac{1}{\sqrt{r_{1i}^2 + r_{1j}^2 - 2r_{1i}r_{1j}[A_{ij} + B_{ij} + C_{ij}]}}.
\end{aligned} \tag{4.16}
$$

We need to calculate derivatives in accordance with equation (4.14):

$$\frac{\partial\phi}{\partial r_{1k}} = -\frac{1}{r_{1k}^2} + \sum_{i=3}^{N-1}\sum_{j=i+1}^{N} \frac{(r_{1i} - r_{1i}(A_{ij}+B_{ij}+C_{ij}))\delta_{ik} + r_{1i}r_{1j}\left(\frac{\partial A_{ij}}{\partial r_{1k}} + \frac{\partial B_{ij}}{\partial r_{1k}} + \frac{\partial C_{ij}}{\partial r_{1k}}\right)}{\left(r_{1i}^2 + r_{1j}^2 - 2r_{1i}r_{1j}[A_{ij}+B_{ij}+C_{ij}]\right)^{3/2}},$$

$$= -\frac{1}{r_{1k}^2} + \mathbb{F}(\mathbf{r}), \tag{4.17}$$

$$\frac{\partial\phi}{\partial r_{2k}} = -\frac{1}{r_{2k}^2} + \sum_{i=3}^{N-1}\sum_{j=i+1}^{N} \frac{r_{1i}r_{1j}\left(\frac{\partial A_{ij}}{\partial r_{2k}} + \frac{\partial B_{ij}}{\partial r_{2k}} + \frac{\partial C_{ij}}{\partial r_{2k}}\right)}{\left(r_{1i}^2 + r_{1j}^2 - 2r_{1i}r_{1j}[A_{ij}+B_{ij}+C_{ij}]\right)^{3/2}},$$

$$= -\frac{1}{r_{2k}^2} + \mathbb{G}(\mathbf{r}), \tag{4.18}$$

$$\frac{\partial\phi}{\partial r_{3k}} = \sum_{i=3}^{N-1}\sum_{j=i+1}^{N} \frac{r_{1i}r_{1j}\left(\frac{\partial A_{ij}}{\partial r_{3k}} + \frac{\partial B_{ij}}{\partial r_{3k}}\right)}{\left(r_{1i}^2 + r_{1j}^2 - 2r_{1i}r_{1j}[A_{ij}+B_{ij}+C_{ij}]\right)^{3/2}},$$

$$= -\frac{1}{r_{3k}^2} + \mathbb{H}(\mathbf{r}) \tag{4.19}$$

where $k > 1$ in equation (4.17), $k > 2$ in equation (4.18) and $k > 3$ in equation (4.19). We note that the first term in equations (4.17) and (4.18) are exactly what one would have with the position dependent parameterisation of $U(\mathbf{q})$ where $\partial U/\partial r_{ij} = -1/r_{ij}^2$. However, at the cost of the linear scaling of force contributions in equation (4.14); we pick up the second term corresponding to the fact we have distance-distance dependencies. The derivatives of A_{ij}, B_{ij}, and C_{ij} are not hard to calculate though computationally they will contribute more flops to the force calculation process than the equivalent derivative calculation in the case the minimal distance description isn't used. The number of derivatives is drastically reduced and scales as $O(N)$, however in the worst case, evaluation of functions $\mathbb{F}, \mathbb{G}$ and $\mathbb{H}$ scale as $O(N^2)$ due to the complexity of ϕ increasing.

4.7 Conclusion

Theorems 1 and 2 show that symmetries of the many-body system imply that the potential U can be written in a form which only depends on pairwise distances between atoms. Since U is a function of the $3N$-dimensional state space, Theorem 1 provides a non-linear transformation of U from a function of $3N$ variables $\mathbf{q}$ into a function of $N(N-1)/2$ distance variables, which is not optimal as it is shown in Theorem 2 where the number of distance variables scales linearly with N.

Considering the example potential (4.3), which depends on all $N(N-1)/2$ distance variables, Theorem 2 provides a reduction of the number of distance variables to $O(N)$. However, if we use the resulting form of the potential, ϕ, this does not directly translate to $O(N)$ complexity of evaluating ϕ. We illustrated this with our example potential (4.3) with $\Psi_2(r_{ij}) = 1/r_{ij}$ and $\Psi_3 \equiv 0$, giving

$$U(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N) = \sum_{\{i,j\}\in\mathcal{S}} \frac{1}{r_{ij}} + \sum_{\{i,j\}\notin\mathcal{S}} \frac{1}{r_{ij}}, \tag{4.20}$$

where r_{ij} is defined by (4.2) and $\mathcal{S}$ is the set of pairs of indices $\{i, j\}$ corresponding to the subset of distances which is used to define ϕ in Theorem 2.

Theorem 2 shows that the number of elements in the set $\mathcal{S}$ scales as $O(N)$, i.e. the number of terms in the first sum on the right hand side of (4.20) is $O(N)$, while the number of terms in the second sum on the right hand side of (4.20) scales as $O(N^2)$. Considering $\{i, j\} \notin \mathcal{S}$, we can find $k \in \{1, 2, \ldots, N\}$ such that $\{i, k\} \in \mathcal{S}$ and $\{j, k\} \in \mathcal{S}$. In particular, distance r_{ij} for $\{i, j\} \notin \mathcal{S}$ can be expressed in terms of distances r_{ik} and r_{jk} using the cosine rule. Therefore, we can find an explicit form of the potential U as a function of $O(N)$ distances corresponding to the indices in the set $\mathcal{S}$. However, the second term in the form (4.20) will contain summations over $O(N^2)$ terms. That is, Theorem 2 does not reduce the $O(N^2)$ complexity of calculations of ϕ. It has been included to illustrate that the number of distance variables needed

scales linearly with N in the same way as the dimension of the phase space scales linearly with N.

Theorem 1 has been formulated as an implication, stating that symmetries (i)-(iv) of a central potential function in Definition 1 imply that the potential can be written as a function of pairwise distances (4.2). However, translations, rotations and reflections are Euclidean isometries, preserving pairwise distances between atoms, so a partial inverse of Theorem 1 also holds, i.e. any potential given as a function of pairwise distances satisfies symmetry assumptions (i)-(iii). The property (iv) states that we consider systems of identical atoms in this chapter. In particular, symmetries (i)-(iv) are both necessary and sufficient conditions for a potential to be expressed as a function of pairwise distances for systems of identical atoms.

We can generalize Theorems 1 and 2 to mixtures of atoms, i.e. for systems when symmetry (iv) in Definition 1 does not hold. Then properties (i)-(iii) of potential U imply that it can be expressed as a function of pairwise distances. If we further reduce the number of symmetries the potential U has, then we can find potential functions which cannot be expressed as a function of pairwise distances. For example,

$$U(\mathbf{q}) = U(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N) = \mathbf{d} \cdot (\mathbf{q}_2 - \mathbf{q}_1)$$

for any nonzero constant vector $\mathbf{d}$ satisfies the translational symmetry (i), but not the rotational symmetry (ii). An example of a potential function satisfying the rotational symmetry (ii) but not the translational symmetry (i) is $U(\mathbf{q}) = |\mathbf{q_1}|$. In fact, symmetries (i)-(iii) are both necessary and sufficient conditions for a potential to be expressed as a function of pairwise distances (for systems of non-identical atoms). Finally, a central potential U gives rise to reciprocal pairwise forces, therefore symmetries (i)-(iii) can be systematically broken to obtain a potential function that governs non-reciprocal interactions.

Chapter 5

Conclusion

The aim of this work can be summarised as improving the understanding of forces involved in many-body systems. Force distributions have been investigated directly, in Chapter 2 and Chapter 3, or by focusing on developing the theory of interatomic potentials in Chapter 4. The work on force distributions largely explored standardised moments, in the low temperature and small number density limit for 1d and 3d simple fluids, analytically and by using MD. Force inversion was then introduced and an approach to achieve this numerically was investigated. The final work addressed when it is possible and practical for pairwise distances to characterise forces as opposed to the full configuration.

In Chapter 2, we derive an expression for the standardised moments of force in (2.8) which allows us to investigate the 1d case via direct variations in box width and temperature. This method matches up well with integral approximations where number density dependence is derived (2.17), and temperature dependence is given in (2.31). We also show that the 1d partition function can be written entirely in terms of standardised moments of force in (2.2). Many-body systems in 3d are studied analytically for $N = 2$ and with MD for $N > 2$. The power law scaling of number density dependence in (2.39) is supported by MD, and clustering mechanics are observed.

We begin Chapter 3 by developing the foundations of force inversion and addressing the harmonic potential in unbounded and also cuboidal domains. In the case of PBCs, we show that inversion typically produced high density configurations, though a lattice transformation can be introduced to spread the configuration out. Inverting non-linear LJ forces is trickier and we develop a NHCM to aid the inversion (3.18); which relies on going from purely harmonic forces to purely LJ forces. Algorithm 1 provided a good start and we explored issues of convergence, implementing the modifications of fine-tuning, and re-looping in Algorithm 2. Direct inversion of the $N = 2$ case is explored and Algorithm 3 is developed which pieces a continuous trajectory together from force data.

Finally, in Chapter 4, we ask what symmetries are sufficient for us to conclude when interactions of many-body systems can be described by pairwise distances. Central potentials are defined as functions exhibiting some translational, rotational, reflectional and parity symmetry. Theorem 1 concludes that these potentials can be written in terms of all pairwise distances; Theorem 2 states we only need a proper subset of those. The minimal set of distances is explicitly constructed; an interesting by-product is the proposal of a new gift-wrapping algorithm. We discuss the limitations of writing a potential with this set of minimal distances, using the gravitational potential as a pedagogical example.

5.1 Further extensions to my work

We now highlight where and how key improvements can be made to various aspects of work in Chapters 2, 3 and 4. We also provide possible future applications as an idea of where we see the work heading.

The aim of Chapter 2 was to give an analytic description of how the standardised moments of force relied on number density and temperature; a feat achieved in 1d and partially in 3d. Temperature dependence was hard to investigate analytically

with the integral approximation methods we utilised. This was due to the fact that Laplace's method greatly struggles on rugged potential energy landscapes admitted by 3d many-body systems [116, 224]. One potential direction would be to use mean field methods to get a better understanding of how temperature plays a role. Such mean field methods have been used in the context of MD for modelling bilayers [135] and protein-solvent systems [221].

Although we provided treatment to LJ fluids, the far-field integral approximation and Laplace calculations (specifically expanding about the potential minimum) can be applied to any potential that has a decaying power law scaling such as the Buckingham [30], Yukawa [236] and Coulombic potentials, so work could be extended to include these. Finally, mathematical techniques used to reconstruct the entire distribution from knowledge of a finite number of moments could be explored with application to the force distribution [127], especially in the context of low density limits where we have an explicit form of the power scaling.

The NHCM from [95] used in Chapter 3 was chosen due to its simplicity. Modifications have made the method more convergent, though this is only practical for up to 20 atoms whereas in theory we want to be able to solve for a system of any number of atoms with the proposed method. One possible improvement is to develop criterion that suggests whether a configuration will work as an initial guess for the given algorithms. The use of machine learning could play an important role here: one could provide a data set of initial configurations that lead to either divergence or convergence, and insight could be given in to how to develop a robust algorithm for force inversion using NHCM [58, 105]. An alternative approach is to implement and adapt another NHCM with reported global convergence such as Sun's [208]. We could potentially improve our own NHCM by using arc length continuation in areas near a singular Jacobian [134] or by using quasi-Newton's methods [17] to avoid the algorithm producing configurations that converge to a singular Jacobian.

Future application of force inversion is of particular interest: it can potentially be used to develop Monte-Carlo schemes whereby forces are randomly perturbed directly instead of positions. This would give a way to develop schemes useful in protein modelling applications where energy minima are particularly hard to escape due to configurational folding [146, 114], and a direct change in the force can move a configuration out of this potential well in the energy landscape.

As we have shown the force space and configuration space are equivalent for some potentials; this opens the pathway to implementing MD in the force-momentum space as opposed to the canonical position-momentum space. This may result in reduced complexity by circumventing the need for $O(N^2)$ pairwise force calculations, with the benefit that positions can be obtained at any time via force inversion. One clear limitation of this description is that any gradient calculation required in time stepping the equations of motion, will generally require a force inversion which scales at least as $O(N^2)$ due to matrix inversion. There is potential to develop this idea further if explicit dependence on atomic positions can be avoided.

The results in Chapter 4 are general, however these were motivated by the field of non-reciprocal interactions. Noting the symmetries that give rise to reciprocal interactions, one can systematically create potentials that model non-reciprocal interactions by breaking each symmetry. One possible direction is to investigate the use of non-reciprocal interactions in self-assembly [169] by introducing potentials that tether individual atoms to specific potential wells while maintaining pairwise interactions. In this way, translational and rotational invariance are both broken. Given that a unique equilibrium configuration is obtained when all atoms are positioned in their respective wells, we can implement MD to simulate this system until equilibrium, or self-assembly, is achieved.

This model is described by non-reciprocal interactions and could give rise to interesting dynamics whilst providing a simple yet intuitive description for self-assembly [68,

211]. Natural avenues to explore are investigating how long it takes for a system to assemble given different initial levels of mixing and different initial average distances from each atom to their respective potential well. In addition to this, the well positions themselves can be varied and the relation of 'well density' to assembly time can be studied.

5.2 How my work relates to the field of molecular dynamics

We conclude the thesis by providing slightly broader strokes, painting the picture of where this work fits in to the field of MD.

The main problem MD faces is handling the computational complexity required to model certain features of interacting systems. This was introduced in Section 1.3 and understanding what we want to describe informs what resolution a model can feasibly be. Coarse graining gives us a way to model a system in a more computationally efficient way while still maintaining the essence of what we wish to capture; if that is dynamics, then force-matching works well [166]. The work of Chapter 2 helps analyse the behaviour of force distributions through standardised moments. These moments can be used in conjunction with bottom-up CG force-matching methods like MS-CG [83], or to directly parameterise other CG models [49] to capture the non-Gaussian force distribution seen in simple fluids [198, 32]. This compliments work done in this area based on density functional theory [181] for high density systems with a bounded potential (finite at zero interatomic distance). Instead, we have investigated the low density limit (and small temperature limit) of the force distribution for LJ fluids, which is not bounded.

Chapter 3 largely ties in with the modelling of bio-molecular structures with its applications to bonded interactions. In order to reduce the complexity of the model, bond deformations are viewed as oscillations around an equilibrium configuration - the potential is modelled as a harmonic oscillator where spring constants are derived

from a covariance matrix [14]. The analysis on the inversion of force data to reproduce a harmonic configuration (creating a harmonic system) addressed in Chapter 3 can be use to provide an alternative method for simplifying a system, at any stage, to one governed by harmonic forces. Similarly, the idea of force inversion can be used to simplify a system to one governed by any pairwise potential, though if the forces are non-linear, a NHCM can be used.

One of the main aims of MD is to aid understanding of mechanisms underlying chemical and physical phenomena. Self assembly is one such mechanism that MD has been employed to understand in the field of intrinsically disordered proteins - this is called liquid-liquid phase separation [69, 5, 38]. The work undertaken in Chapter 4 is useful for giving a way to bridge the gap between this field, and non-reciprocal interactions, as models can be developed by breaking symmetries that induce self assembly, as mentioned in Section 5.1.

The work presented in this thesis has furthered the theoretical understanding of forces in general, providing novel insights which have far reaching applications in MD.

Bibliography

[1] S. Abbasbandy, Y. Tan, and S.J. Liao. Newton-homotopy analysis method for nonlinear equations. *Applied Mathematics and Computation*, 188(2):1794–1800, 2007.

[2] G.J. Ackland. 1.10 - interatomic potential development. In *Comprehensive Nuclear Materials*, pages 267–291. Elsevier, 2012.

[3] A.S. Adcock and A.J. McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical Reviews*, 106(5):1589–1615, 2006.

[4] N.I. Akhiezer. *The Classical Moment Problem and Some Related Questions in Analysis*. 2020.

[5] S. Alberti, A. Gladfelter, and T. Mittag. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell*, 176(3):419–434, 2019.

[6] M.P. Allen and D. Quigley. Some comments on Monte Carlo and molecular dynamics methods. *Molecular Physics*, 111(22–23):3442–3447, 2013.

[7] M.P. Allen and D.J. Tildesley. *Computer Simulation of Liquids*, pages 185–215. Oxford Academic, 2017.

[8] E.L. Allgower and K. Georg. *Introduction to Numerical Continuation Methods*. Society for Industrial and Applied Mathematics, 2003.

[9] F.J. Angeles and K.J. Harmon. Nonreciprocal interactions induced by water in confinement. *Physical review Res.*, 2, 2020.

[10] S.J. Antony. Evolution of force distribution in three-dimensional granular media. *Physical review E*, 63, 2001.

[11] P.R. Arantes, A. Saha, and G. Palermo. Fighting COVID-19 using molecular dynamics simulations. *ACS Central Science*, 6(10):1654–1656, 2020.

[12] O. Axelsson and S. Sysala. Continuation Newton methods. *Computers and Mathematics with Applications*, 70(11):2621–2637, 2015.

[13] K. Bagi. Statistical analysis of contact force components in random granular assemblies. *Granular Matter*, 5:45–54, 2003.

[14] I. Bahar, T.R. Lezon, A. Bakan, and I.H. Shrivastava. Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem Rev*, 110(3):1463–1497, 2010.

[15] V.G. Baidakov, G.G. Chernykh, and S.P. Protsenko. Effect of the cut-off radius of the intermolecular potential on phase equilibrium and surface tension in Lennard–Jones systems. *Chemical Physics Letters*, 321(3):315–320, 2000.

[16] C.M. Baker. Polarizable force fields for molecular dynamics simulations of biomolecules. *WIREs Computational Molecular Science*, 5(2):241–254, 2015.

[17] R.E. Bank and D.J. Rose. Global approximate Newton methods. *Numerische Mathematik*, 37:279–295, 1981.

[18] J.H. Barnes and P. Hut. A hierarchical $O(N \log(N))$ force-calculation algorithm. *Nature*, 324:446–449, 2022.

[19] V. Beato, L. Pietronero, and S. Zapperi. Statistical properties of dislocation mutual interactions. *The Journal of Statistical Mechanics: Theory and Experiment*, 2005, 2005.

[20] C.M. Bender and S.A. Orszag. *Advanced mathematical methods for scientists and engineers: asymptotic methods and perturbation theory: v. 1.* Springer, 1999.

[21] H.J.C. Berendsen. *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*, pages 3–18. Cambridge University Press, 2007.

[22] M. Berezowski and M. Lawnik. Homotopic parametric continuation method for determining stationary states of chemical reactors with dispersion. *Symmetry*, 13(12), 2021.

[23] R. Biswas and D.R. Hamann. New classical models for silicon structural energies. *Physical Review B*, 36:6434–6445, 1987.

[24] J-M. Bomont and J-L. Bretonnet. An effective pair potential for thermodynamics and structural properties of liquid mercury. *The Journal of Chemical Physics*, 124(5), 2006.

[25] M. Bonitz, C. Henning, and D. Block. Complex plasmas: a laboratory for strong correlations. *Reports on Progress in Physics*, 73(6), 2010.

[26] J.P. Boon and S. Yip. *Molecular hydrodynamics*. McGraw-Hill New York ; London, 1980.

[27] F.H. Branin. Widely convergent method for finding multiple solutions of simultaneous nonlinear equations. *The IBM Journal of Research and Development*, 16(5):504–522, 1972.

[28] A.C. Brańka, D.M. Heyes, and G. Rickayzen. Pair force distributions in simple fluids. *The Journal of Chemical Physics*, 135(164507), 2011.

[29] R.C. Büchel, D.A. Rudolph, and I. Frank. Deterministic quantum mechanics: The role of the Maxwell–Boltzmann distribution. *Int J Quantum Chem*, 121, 2021.

[30] R.A. Buckingham and J.E. Jones. The classical equation of state of gaseous helium, neon and argon. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 168(933):264–283, 1938.

[31] Z. Cang, L. Mu, and G-W Wei. Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening. *PLoS Comput Biol*, 14(1), 2018.

[32] A. Carof, R. Vuilleumier, and B. Rotenberg. Two algorithms to compute projected correlation functions in molecular dynamics simulations. *JThe Journal of Chemical Physics*, 140(124103), 2014.

[33] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Review of Modern Physics*, 15(1), 1943.

[34] P.H. Chavanis. *Statistical Mechanics of Two-Dimensional Vortices and Stellar Systems*, pages 208–289. Springer Berlin Heidelberg, 2002.

[35] J.T. Chayes and L. Chayes. On the validity of the inverse conjecture in classical density functional theory. *The Journal of Statistical Physics*, 36:471—-488, 1984.

[36] J.C. Chen and A.S. Kim. Brownian dynamics, Molecular Dynamics, and Monte Carlo modeling of colloidal systems. *Advances in Colloid and Interface Science*, 112(1):159–173, 2004.

[37] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An nlog(n) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[38] S. Das, A. Amin, Y-H. Lin, and H.S. Chan. Coarse-grained residue-based models of disordered protein condensates: utility and limitations of simple charge pattern parameters. *The Journal of Physical Chemistry Chemical Physics*, 20:28558–28574, 2018.

[39] A. Davtyan, J.F Dama, G.A. Voth, and H.C. Andersen. Dynamic force matching: A method for constructing dynamical coarse-grained models with realistic time dependence. *The Journal of Chemical Physics*, 142, 2015.

[40] A. Davtyan, G.A. Voth, and H.C. Andersen. Dynamic force matching: Construction of dynamic coarse-grained models with realistic short time dynamics and accurate long time dynamics. *The Journal of Chemical Physics*, 145(22), 2016.

[41] M.S. Daw and M.I. Baskes. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Physical Review B*, 29:6443–6453, 1984.

[42] F.A. Garcia Daza, A.J. Colville, and A.D. Mackie. Mean-field coarse-grained model for poly(ethylene oxide)-poly(propylene oxide)-poly(ethylene oxide) triblock copolymer systems. *Langmuir*, 31(12):0743–7463, 2015.

[43] L. DeCarlo. On the meaning and use of kurtosis. *Psychological Methods*, 2(3):292–307, 1997.

[44] R. Drautz, M. Fähnle, and J.M. Sanchez. General relations between many-body potentials and cluster expansions in multicomponent systems. *The Journal of Physics: Condensed Matter*, 16(23), 2004.

[45] N.J. Dunn, T.T. Foley, and W.G. Noid. Van der Waals perspective on coarse-graining: Progress toward solving representability and transferability problems. *Accounts of Chemical Research*, 49(1):2832–2840, 2016.

[46] D. Dürr, S. Goldstein, and J.L. Lebowitz. A mechanical model of Brownian motion. *Commun. Math. Phys.*, 78:507–530, 1981.

[47] R. Erban. From molecular dynamics to Brownian dynamics. *Proceedings of the Royal Society A*, 470(2167), 2014.

[48] R. Erban. Coupling all-atom molecular dynamics simulations of ions in water with Brownian dynamics. *Proceedings of the Royal Society A*, 472(2186), 2016.

[49] R. Erban. Coarse-graining molecular dynamics: stochastic models with non-Gaussian force distributions. *The Journal of Mathematical Biology*, 80:457–479, 2020.

[50] R. Erban and S. J. Chapman. *Stochastic modelling of reaction-diffusion processes*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2020.

[51] F. Ercolessi and J.B. Adams. Interatomic potentials from first-principles calculations: The force-matching method. *Europhysics Letters*, 26(8):583, 1994.

[52] P. Español and I. Zúñiga. Force autocorrelation function in Brownian motion theory. *The Journal of Chemical Physics*, 98(1):574–580, 1993.

[53] A. Gabrielli et al. Force distribution in a randomly perturbed lattice of identical atoms with $1/r^2$ pair interaction. *Physical Review E*, 74(021110), 2006.

[54] A. George et al. Review of electrostatic force calculation methods and their acceleration in molecular dynamics packages using graphics processors. *ACS Omega*, 7(37):32877–32896, 2022.

[55] A. Ramírez-Hernández et al. Demixing by a nematic mean field: Coarse-grained simulations of liquid crystalline polymers. *Polymers*, 9, 2017.

[56] A. Sen et al. Chemo and phototactic nano/microbots. *Faraday Discuss.*, 143:15–27, 2009.

[57] A.P. Lyubartsev et al. Systematic hierarchical coarse-graining with the inverse Monte Carlo method. *The Journal of Chemical Physics*, 143(24), 2015.

[58] A.S. Berahas et al. Quasi-Newton methods for machine learning: forget the past, just sample. *Optimization Methods and Software*, 37(5):1668–1704, 2022.

[59] A.V. Ivlev et al. Complex plasmas in external fields: The role of non-Hamiltonian interactions. *Physical review Letters*, 106:155001, 2011.

[60] C. Ayaz et al. Generalized Langevin equation with a nonlinear potential of mean force and nonlinear memory friction from a hybrid projection scheme. *Physical review E*, 105, 2022.

[61] C. Zhao et al. Seven-site effective pair potential for simulating liquid water. *Physical Chemistry B*, 123(21):4594–4603, 2019.

[62] F. Kümmel et al. Circular motion of asymmetric self-propelling particles. *Physical review Letters*, 110:198302, 2013.

[63] F. Mouvet et al. Recent advances in first-principles based molecular dynamics. *Accounts of Chemical Research*, 55(3):221–230, 2022.

[64] F.R. Souza et al. Recent open issues in coarse grained force fields. *The Journal of Chemical Information and Modeling*, 60(12):5881–5884, 2020.

[65] G.A. Cisneros et al. Modeling molecular interactions in water: From pairwise to many-body potential energy functions. *Chemical Reviews*, 116(13):7501–7528, 2016.

[66] G.C. Pruteanu et al. Krypton and the fundamental flaw of the Lennard-Jones potential. *The Journal of Physical Chemistry Letters*, 13(35):8284–8289, 2022.

[67] H. Ingólfsson et al. The power of coarse graining in biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(3):225–248, 2013.

[68] J. Hautman et al. Molecular dynamics investigations of self-assembled monolayers. *J. Chem. Soc., Faraday Trans.*, 87:2031–2037, 1991.

[69] J. McCarty et al. Complete phase diagram for liquid–liquid phase separation of intrinsically disordered proteins. *The Journal of Physical Chemistry Letters*, 10(8):1644–1652, 2019.

[70] J. Wang et al. Machine learning of coarse-grained molecular dynamics force fields. *ACS Central Science*, 5(5):755–767, 2019.

[71] J.F. Dama et al. The theory of ultra-coarse-graining. 1. general principles. *The Journal of Chemical Theory and Computation*, 9(5):2466–2480, 2013.

[72] L. Delle Site et al. Molecular dynamics of open systems: Construction of a mean-field particle reservoir. *Advanced Theory and Simulations*, 2(5), 2019.

[73] M. Fitzner et al. Communication: Truncated non-bonded potentials can yield unphysical behavior in molecular dynamics simulations of interfaces. *The Journal of Chemical Physics*, 147(12), 2017.

[74] M.G. Guenza et al. Accuracy, transferability, and efficiency of coarse-grained models of molecular liquids. *The Journal of Chemical Physics*, 122(45):10257—-10278, 2018.

[75] P. Pattnaik et al. Machine learning for accurate force calculations in molecular dynamics simulations. *The Journal of Physical Chemistry A*, 124(34):6954–6967, 2020.

[76] S. Duane et al. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987.

[77] S. Kmiecik et al. Coarse-grained protein models and their applications. *Chemical Reviews*, 116(14):7898–7936, 2016.

[78] S. Reiber et al. Real cost of speed: The effect of a time-saving multiple-time-stepping algorithm on the accuracy of molecular dynamics simulations. *The Journal of Chemical Theory and Computation*, 13(6):2367–2372, 2017.

[79] S.A.M. Loos et al. Nonreciprocal forces enable cold-to-hot heat transfer between nanoparticles. *Sci Rep*, 13(4517), 2023.

[80] S.E. Feller et al. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*, 103(11):4613–4621, 1995.

[81] S.J. Marrink et al. The martini force field: coarse grained model for biomolecular simulations. *Physical Chemistry B*, 111(27):7812–7824, 2007.

[82] W. Tschöp et al. Simulation of polymer melts. i. coarse-graining procedure for polycarbonates. *Acta Polymerica*, 49(2-3):61–74, 1998.

[83] W.G. Noid et al. The multiscale coarse-graining method. i. a rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics*, 128, 2008.

[84] X. Wang et al. The Lennard-Jones potential: when (not) to use it. *The Journal of Physical Chemistry Chemical Physics*, 22:10624–10633, 2020.

[85] Z. Xia et al. Coarse-grained model for simulation of rna three-dimensional structures. *The Journal of Chemical Physics*, 114(42):13497—-13506, 2010.

[86] R. Evans. Comment on reverse Monte Carlo simulation. *Molecular Simulation*, 4(6):409–411, 1990.

[87] G. Favrin, A. Irbäck, and S. Wallin. Folding of a small helical protein using hydrogen bonds and hydrophobicity forces. *Proteins: Structure, Function, and Bioinformatics*, 47(2):99–105, 2002.

[88] J. Frank and G.A. Gottwald. The Langevin limit of the Nosé-Hoover-Langevin thermostat. *The Journal of Statistical Physics*, 143(4):715–724, 2011.

[89] I. Fredholm. Sur une classe d'équations fonctionnelles. *Acta Mathematica*, 27:365 – 390, 1903.

[90] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*. Cambridge Texts in Applied Mathematics. Academic Press, 2nd edition, 1996.

[91] M. Fruchart, R. Hanai, and P.B. Littlewood. Non-reciprocal phase transitions. *Nature*, 592:363–369, 2021.

[92] J. Gallardo-Alvarado. An application of the Newton-homotopy continuation method for solving the forward kinematic problem of the 3-rrs parallel manipulator. *Mathematical Problems in Engineering*, 2019.

[93] A. Galántai. The theory of Newton's method. *The Journal of Computational and Applied Mathematics*, 124(1):25–44, 2000.

[94] C.B. Garcia and F.J. Gould. Relations between several path following algorithms and local and global Newton methods. *SIAM Review*, 22(3):263–274, 1980.

[95] C.B. Garcia and W.I. Zangwill. Finding all solutions to polynomial systems and other systems of equations. *Mathematical Programming*, 16(1):159–176, 1979.

[96] T.E. Gartner and A. Jayaraman. Modeling and simulations of polymers: A roadmap. *Macromolecules*, 52(3):755–786, 2019.

[97] J.G. Gay and B.J. Berne. Modification of the overlap potential to mimic a linear site–site potential. *The Journal of Chemical Physics*, 74(6):3316–3319, 1981.

[98] R. Ghanati and M.M. Petke. A homotopy continuation inversion of geoelectrical sounding data. *The Journal of Applied Geophysics*, 191(104356), 2021.

[99] D.T. Gillespie. Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. *Physical review E*, 54:2084–2091, 1996.

[100] A. Giró, E. Guardia, and J.A. Padró. Langevin and molecular dynamics simulations of Lennard-Jones liquids. *The Journal of Molecular Physics*, 55(5):1063–1074, 1985.

[101] M.D.L.L. González. A novel predictive homotopic path tracking algorithm to solve non-linear algebraic equations. *The Canadian Journal of Chemical Engineering*, 2022.

[102] L. Greengard. The numerical solution of the N-body problem. *Computers in Physics*, 4(2):142–152, 1990.

[103] G. Grochola, S.P. Russo, and I.K. Snook. On fitting a gold embedded atom method potential using the force matching method. *The Journal of Chemical Physics*, 123, 2005.

[104] R. Gunaratne, D.B Wilson, M.G. Flegg, and R. Erban. On short-range and long-range interactions in multi-resolution dimer models. *Interface Focus*, 9(3), 2019.

[105] T-D. Guo, Y. Liu, and C-Y Han. An overview of stochastic quasi-Newton methods for large-scale machine learning. *The Journal of the Operations Research Society of China*, pages 1–31, 2023.

[106] E. Hairer, C. Lubich, and G. Wanner. Geometric numerical integration illustrated by the Störmer–Verlet method. *Acta Numerica*, 12:399–450, 2003.

[107] K.R. Haldey and C. McCabe. Coarse-grained molecular models of water: A review. *Mol Simul*, 38(8-9):671–681, 2012.

[108] H.W. Hammer, A. Nogga, and A. Schwenk. Colloquium: Three-body forces: From cold atoms to nuclei. *Review of Modern Physics*, 85:197–217, 2013.

[109] J-P. Hansen and L. Verlet. Phase transitions of the Lennard-Jones system. *Physical Review*, 184(151), 1969.

[110] J.P. Hansen and I.R. McDonald. Chapter 11 - Molecular liquids. In *Theory of Simple Liquids (Fourth Edition)*, pages 455–510. 2013.

[111] J.P. Hansen and I.R. McDonald. Chapter 7 - time-dependent correlation and response functions. In *Theory of Simple Liquids (Fourth Edition)*, pages 265–310. 2013.

[112] M.S Harish and P.K Patra. Temperature and its control in molecular dynamics simulations. *Molecular Simulation*, 47(9):701–729, 2021.

[113] S. Harris. Force correlation function representation for the self-diffusion coefficient. *Molecular Physics*, 23(5):861–865, 1972.

[114] T. Heijden and C. Dekker. Monte Carlo simulations of protein assembly, disassembly, and linear motion on dna. *Biophysics J*, 95(10):4560–4569, 2008.

[115] D.R. Herschbach. Molecular dynamics of elementary chemical reactions. *Angew. Chem. Int.*, 26:1221–1243, 1987.

[116] R.D. Hills, L. Lu, and G.A. Voth. Multiscale coarse-graining of the protein energy landscape. *PLOS Computational Biology*, 6:1–12, 2010.

[117] M.W. Hirsch. A proof of the nonretractibility of a cell onto its boundary. *Proceedings of the American Mathematical Society*, 14(2):364–365, 1963.

[118] M.A. Hoef and P.A. Madden. Three-body dispersion contributions to the thermodynamic properties and effective pair interactions in liquid argon. *The Journal of Chemical Physics*, 111(4):1520–1526, 1999.

[119] S.A. Hollingsworth and R.O. Dror. Molecular dynamics simulation for all. *Neuron*, 99(6):1129–1143, 2018.

[120] D.M. Huang and H.T.L. Nguyen. Systematic bottom-up molecular coarse-graining via force and torque matching using anisotropic particles. *The Journal of Chemical Physics*, 156(18), 2022.

[121] W. Humphrey, A. Dalke, and K. Schulten. Vmd - visual molecular dynamics. *The Journal of Molecular Graphics*, 14:33–38, 1996.

[122] P.H. Hünenberger. Thermostat algorithms for molecular dynamics simulations. In C. Holm and K. Kremer, editors, *Advanced Computer Simulation Approaches for Soft Matter Sciences I*, volume 173 of *Advances in Polymer Science*, pages 105–149. Springer, 2005.

[123] B. Isralewitz, M. Gao, and K. Schulten. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology*, 11(2):224–230, 2001.

[124] A.V. Ivlev, J. Bartnick, and M. Heinen et al. Statistical mechanics where Newton's third law is broken. *Physical review X*, 5, 2015.

[125] S. Izvekov and G.A. Voth. A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry B*, 109(7):2469–2473, 2005.

[126] J. Jin, A.J. Pak, and G.A. Voth. Understanding missing entropy in coarse-grained systems: Addressing issues of representability and transferability. *Physical Chemistry Letters*, 10(16):4549–4557, 2019.

[127] V. John, I. Angelov, A.A. Öncül, and D. Thévenin. Techniques for the reconstruction of a distribution from a finite number of its moments. *Chemical Engineering Science*, 62(11):2890–2904, 2007.

[128] J. Jones. On the determination of molecular fields. — ii. from the equation of state of a gas. *Proceedings of the Royal Society A*, 106(738), 1924.

[129] S.Y. Joshi and S.A. Deshmukh. A review of advancements in coarse-grained molecular dynamics simulations. *Molecular Simulation*, 47(10-11):786–803, 2021.

[130] J. Jung, J. Lee, and J. Kim. Cluster growth mechanisms in Lennard-Jones fluids: a comparison between molecular dynamics and Brownian dynamics simulations. *Chemical Physics*, 449:1–9, 2015.

[131] M. Karplus and M. McCammon. Molecular dynamics simulations of biomolecules. *Nat Struct Mol Biol*, 9:646–652, 2002.

[132] A.J. Katan, M.H. van Es, and T.H. Oosterkamp. Quantitative force versus distance measurements in amplitude modulation afm: a novel force inversion technique. *Nanotechnology*, 20(16), 2009.

[133] E.R. Kay, D.A. Leigh, and F. Zerbetto. Synthetic molecular motors and mechanical machines. *Angewandte Chemie International Edition*, 46:72–191, 2007.

[134] H.B. Keller. Global homotopies and Newton methods. In C. de Boor and G.H. Golub, editors, *Recent Advances in Numerical Analysis*, pages 73–94. 1978.

[135] G.A. Khelashvili, S.A. Pandit, and H.L. Scott. Self-consistent mean-field model based on molecular dynamics: application to lipid-cholesterol bilayers. *The Journal of Chemical Physics*, 123(3):263–274, 2005.

[136] D.B. Kinghorn and L. Adamowicz. A new N-body potential and basis set for adiabatic and non-adiabatic variational energy calculations. *The Journal of Chemical Physics*, 106(21):8760–8768, 1997.

[137] S. Kirkpatrick, C. Gelatt Jr, and M. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[138] J. Kohler and F. Noé et al. Force-matching coarse-graining without forces. *arXiv*, 2022.

[139] T.D. Kühne. Second generation Car–Parrinello molecular dynamics. *WIREs Computational Molecular Science*, 4(4):391–406, 2014.

[140] L.D. Landau and E.M. Lifshitz. *Mechanics, Third Edition: Volume 1 (Course of Theoretical Physics)*. Butterworth-Heinemann, 3 edition, 1976.

[141] R. Lazim, D. Suh, and S. Choi. Advances in molecular dynamics simulations and enhanced sampling methods for the study of protein systems. *The International Journal of Molecular Sciences*, 21(17), 2020.

[142] J. Lee and H. Chiang. Convergent regions of the Newton homotopy method for nonlinear systems: theory and computational applications. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 48(1):51–66, 2001.

[143] B. Leimkuhler and C. Matthews. *Molecular dynamics with deterministic and stochastic numerical methods*. Number 39 in Interdisciplinary Applied Mathematics. Springer, 2015.

[144] B.J. Leimkuhler, S. Reich, and R.D. Skeel. *Integration Methods for Molecular Dynamics*, pages 161–185. Springer New York, 1996.

[145] T.Y. Li. Numerical solution of multivariate polynomial systems by homotopy continuation methods. *Acta Numerica*, 6:399—-436, 1997.

[146] Z. Li and H.A. Scheraga. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc Natl Acad Sci U S A*, 84(19):6611—-6615, 1987.

[147] D. Lide. Properties of the elements and inorganic compounds; melting, boiling, triple, and critical temperatures of the elements. In *CRC Handbook of Chemistry and Physics*, chapter 4. CRC Press, 86th edition, 2005.

[148] I.I. Lisina and O.S. Vaulina. The effect of nonreciprocal interaction on the redistribution of kinetic energy in the system of particles. *J. Phys.: Conf. Ser*, 653, 2015.

[149] A. Liwo and C. Czaplewski. Extension of the force-matching method to coarse-grained models with axially symmetric sites to produce transferable force fields: Application to the UNRES model of proteins. *The Journal of chemical physics*, 152(5), 2020.

[150] S. Loos and S. Klapp. Irreversibility, heat and information flows induced by non-reciprocal interactions. *The New Journal of Physics*, 22(12), 2020.

[151] W. Loose and S. Hess. Nonequilibrium velocity distribution function of gases: Kinetic theory and molecular dynamics. *Physical review A*, 37:2099–2111, 1988.

[152] A.A. Louis. Beware of density dependent pair potentials. *The Journal of Physics: Condensed Matter*, 14(40), 2002.

[153] P.A. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Physical review E*, 52:3730–3737, 1995.

[154] H.M. Marques and K.L. Brown. Molecular mechanics and molecular dynamics simulations of porphyrins, metalloporphyrins, heme proteins and cobalt corrinoids. *Coordination Chemistry Reviews*, 225(1):123–158, 2002.

[155] T.W. Marshall. The approach to the Boltzmann distribution in Brownian motion. *Z. angew. Math. Phys.*, 34:241—-245, 1983.

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

[157] E. Mastny and J. de Pablo. Melting line of the Lennard-Jones system, infinite size, and full potential. *The Journal of Chemical Physics*, 127(104504), 2007.

[158] J.C. Maxwell. Illustrations of the dynamical theory of gases.—part i. on the motions and collisions of perfectly elastic spheres. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, pages 19–32, 1860.

[159] R.I. McLachlan and P. Atela. The accuracy of symplectic integrators. *Nonlinearity*, 5(2), 1992.

[160] C.C. Moore. Ergodic theorem, ergodic theory, and statistical mechanics. *Proc Natl Acad Sci U S A*, 112(7):1907–1911, 2015.

[161] A. Morriss-Andrews and J-E Shea. Computational studies of protein aggregation: Methods and applications. *Annual Review of Physical Chemistry*, 66(1):643–666, 2015.

[162] P.M. Morse. Diatomic molecules according to the wave mechanics. ii. vibrational levels. *Physical review*, 34:57–64, 1929.

[163] J.N. Murrell. *Molecular potential energy functions*. Wiley, 1984.

[164] S. Nadarajah. Some explicit expressions for the probability distribution of force magnitude. *Sadhana*, 33:357–365, 2008.

[165] P. Nicolini, E. Guàrdia, and M. Masia. Shortcomings of the standard Lennard–Jones dispersion term in water models, studied with force matching. *The Journal of Chemical Physics*, 139(18), 2013.

[166] W.G. Noid. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics*, 139(090901), 2013.

[167] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[168] R. Nussinov. The significance of the 2013 nobel prize in chemistry and the challenges ahead. *PLoS Comput Biol*, 10(1), 2014.

[169] S. Osat and R. Golestanian. Non-reciprocal multifarious self-organization. *Nat. Nanotechnol*, 18:79–85, 2023.

[170] E. Paquet and H.L. Viktor. Molecular dynamics, Monte Carlo simulations, and Langevin dynamics: A computational review. *BioMed Research International*, 2015.

[171] L. Pareschi and G. Russo. An introduction to Monte Carlo method for the Boltzmann equation. *ESAIM: Proc*, 10:35–75, 2001.

[172] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *The Journal of Applied Physics*, 52(12):7182–7190, 1981.

[173] J. Petravic. Force autocorrelation function in linear response theory and the origin of friction. *The Journal of Chemical Physics*, 129(9), 2008.

[174] S.J. Plimpton and A.P. Thompson. Computational aspects of many-body potentials. *MRS Bulletin*, 37(5):513–521, 2012.

[175] A. Del Popolo and M. Gambera. The statistics of the gravitational field arising from an inhomogenous system of particles. *Astronomy Astrophysics*, 342:34–40, 1999.

[176] O.V. Prezhdo and P.J. Rossky. Mean-field molecular dynamics with surface hopping. *The Journal of Chemical Physics*, 107(3):825–834, 1997.

[177] A. Rahman. Correlations in the motion of atoms in liquid argon. *Physical Review A*, 136(2), 1964.

[178] D. Shemesh R.B. Gerber and M. E. Varner. Ab initio and semi-empirical molecular dynamics simulations of chemical reactions in isolated molecules and in clusters. *The Journal of Physical Chemistry Chemical Physics*, 16:9760–9775, 2014.

[179] D. Reith, M. Pütz, and F. Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.*, 24:1624–1636, 2003.

[180] R.M. Richard, K.U. Lao, and J.M. Herbert. Understanding the many-body expansion for large systems. i. precision considerations. *The Journal of Chemical Physics*, 141(1), 2014.

[181] G. Rickayzen, A.C. Brańka, S. Pieprzyk, and D.M. Heyes. Single atom force distributions in simple fluids. *The Journal of Chemical Physics*, 137(094505), 2012.

[182] V. Rokhlin. Rapid solution of integral equations of classical potential theory. *The Journal of Computational Physics*, 60(2):187–207, 1985.

[183] E. Rolls, Y. Togashi, and R. Erban. Varying the resolution of the rouse model on temporal and spatial scales: application to multiscale modelling of DNA dynamics. *Multiscale Modeling and Simulation*, 15(4):1672–1693, 2017.

[184] P.E. Rouse. A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *The Journal of Chemical Physics*, 21(7):1272–1280, 1953.

[185] L. Rowley, D. Nicholson, and N.G. Parsonage. Monte Carlo grand canonical ensemble calculation in a gas-liquid transition region for 12-6 argon. *The Journal of Computational Physics*, 17(4):401–414, 1975.

[186] R.E. Rudd and J.Q. Broughton. Coarse-grained molecular dynamics and the atomic limit of finite elements. *Physical review B*, 58:R5893–R5896, 1998.

[187] C. Sagui and T. Darden. Multigrid methods for classical molecular dynamics simulations of biomolecules. *The Journal of Chemical Physics*, 114(15):6578–6591, 2001.

[188] A.A. Samoletov, C.P. Dettmann, and M.A.J. Chaplain. Thermostats for "slow" configurational modes. *The Journal of Statistical Physics*, 128:1321–1336, 2007.

[189] J.M. Sanchez, F. Ducastelle, and D. Gratias. Generalized cluster description of multicomponent systems. *Physica A: Statistical Mechanics and its Applications*, 128(1):334–350, 1984.

[190] H. Scarf. The approximation of fixed points of a continuous mapping. *SIAM Journal on Applied Mathematics*, 15(5):1328–1343, 1967.

[191] T. Schlick. *Molecular modeling and simulation*. Springer, 2002.

[192] T. Schlick and S. Portiloo-Ledesma. Biomolecular modeling thrives in the age of technology. *Nature Computational Science*, 1(5):321–331, 2021.

[193] A. Schultz and D. Kofke. Erratum: 'comprehensive high-precision high-accuracy equation of state and coexistence properties for classical Lennard-Jones crystals and low-temperature fluid phases'. *The Journal of Chemical Physics*, 153(059901), 2020.

[194] D.J. Searles and D.J. Evans. The fluctuation theorem and Green–Kubo relations. *The Journal of Chemical Physics*, 112(22):9727–9735, 2000.

[195] H.M. Senn and W. Thiel. Qm/mm methods for biomolecular systems. *Angew Chem Int Ed Engl*, 48:1198–1229, 2009.

[196] M.G. Sexton, J.E. Socolar, and D.G. Schaeffer. Force distribution in a scalar model for noncohesive granular material. *Physical review E*, 60:1999–2008, 1999.

[197] L. Shijun. Homotopy analysis method: A new analytical technique for nonlinear problems. *Communications in Nonlinear Science and Numerical Simulation*, 2(2):95–100, 1997.

[198] H.K. Shin, C. Kim, P. Talkner, and E.K. Lee. Brownian motion from molecular dynamics. *Chemical Physics*, 375(2–3):316–326, 2010.

[199] P.K. Shukla and B. Eliasson. Colloquium: Fundamentals of dust-plasma interactions. *Review of Modern Physics*, 81:25–44, 2009.

[200] R.D. Skeel. What makes molecular dynamics work? *SIAM J Sci Comput.*, 31(2):1363–1378, 2009.

[201] R. Soto and R. Golestanian. Self-assembly of catalytically active colloidal molecules: Tailoring activity through surface chemistry. *Physical review Letters*, 112, 2014.

[202] M. Sprik. *Effective Pair Potentials and Beyond*, pages 211–259. Springer Netherlands, 1993.

[203] W. Stacklies and F. Graeter. Force distribution in proteins from molecular dynamics simulations. *AIP Conference Proceedings*, 1102:233–235, 2009.

[204] W. Stacklies, C. Seifert, and F. Graeter. Implementation of force distribution analysis for molecular dynamics simulations. *BMC Bioinformatics*, 12(101):263–274, 2011.

[205] S. Stephan, J. Staubach, and H. Hasse. Review and comparison of equations of state for the Lennard-Jones fluid. *Fluid Phase Equilibria*, 523(112772), 2020.

[206] S. Stephan, M. Thol, J. Vrabec, and H. Hasse. Thermophysical properties of the Lennard-Jones fluid: database and data assessment. *The Journal of Chemical Information and Modeling*, 59(10):4248–4265, 2019.

[207] F.H. Stillinger and T.A. Weber. Computer simulation of local order in condensed phases of silicon. *Physical Review B*, 31:5262–5271, 1985.

[208] A.C. Sun and W.D. Seider. Homotopy-continuation method for stability analysis in the global minimization of the Gibbs free energy. *Fluid Phase Equilibria*, 103(2):213–249, 1995.

[209] K.F. Sundman. Mémoire sur le problème des trois corps. *Acta Mathematica*, 36:105–179, 1913.

[210] J.A. Sánchez, G. Rumi, and R.C. Maldonado. Non-Gaussian tail in the force distribution: a hallmark of correlated disorder in the host media of elastic objects. *Sci Rep*, 10, 2020.

[211] H. Tanaka. General view of a liquid-liquid phase transition. *Physical review E*, 62:6968–6976, 2000.

[212] D.J. Tobias, G.J Martyna, and M.L. Klein. Molecular dynamics simulations of a protein in the canonical ensemble. *The Journal of Physical Chemistry*, 97(49), 1993.

[213] A. Traheem and A. Laradji. Classroom note: a generalization of Leibniz rule for higher derivatives. *The International Journal of Mathematical Education in Science and Technology*, 34(6):905–907, 2003.

[214] M. Tuckerman, B.J. Berne, and G.J. Martyna. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*, 97(3):1990–2001, 1992.

[215] M.E. Tuckerman. Non-Hamiltonian molecular dynamics: generalizing Hamiltonian phase space principles to non-Hamiltonian systems. *The Journal of Chemical Physics*, 115(1678), 2001.

[216] P.F. Tupper. Ergodicity and the numerical simulation of Hamiltonian systems. *SIAM Journal on Applied Dynamical Systems*, 4(3):563–587, 2005.

[217] J. Utterson and R. Erban. On standardised moments of force distribution in simple liquids. *The Journal of Physical Chemistry Chemical Physics*, 24:5646–5657, 2022.

[218] J. Utterson and R. Erban. Symmetries of many-body systems imply distance-dependent potentials. *Physical Review E*, 2023.

[219] T. van Noije and M. Ernst. Velocity distributions in homogeneous granular fluids: the free and the heated case. *Granular Matter*, 1:57–64, 1998.

[220] L. Verlet. Computer 'experiments' on classical fluids. i. thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159(1), 1967.

[221] A.R. Völkel and J. Noolandi. Mean field approach to the thermodynamics of protein–solvent systems with application to p53. *Biophysical Journal*, 80(3):1524–1537, 2001.

[222] G.A. Voth. *Coarse-Graining of Condensed Phase and Biomolecular Systems*, pages 1–4. CRC Press, 2008.

[223] H. Waisman and J. Fish. A space–time multilevel method for molecular dynamics simulations. *Computer Methods in Applied Mechanics and Engineering*, 195(44):6542–6559, 2006.

[224] D. Wales. Exploring energy landscapes. *Annual Review of Physical Chemistry*, 69:401–425, 2018.

[225] Q. Wang. The global solution of the N-body problem. *Celestial Mech Dyn Astr*, 50:73—-88, 1990.

[226] Y. Wang, W.G. Noid, P. Liu, and G.A. Voth. Effective force coarse-graining. *The Journal of Physical Chemistry Chemical Physics*, 11:2001–2015, 2009.

[227] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *The Journal of Molecular Biology*, 103(2):227–249, 1976.

[228] K.O. Waseda. Structure and effective pair potential of liquid silicon. *The Japanese Journal of Applied Physics*, 35(1R), 1996.

[229] L. Wasserman. Topological data analysis. *Annual Review of Statistics and Its Application*, 5(1):501–532, 2018.

[230] H. Watanabe, N. Ito, and C. Hu. Phase diagram and universality of the Lennard-Jones gas-liquid system. *The Journal of Chemical Physics*, 136(204102), 2012.

[231] G. Widmalm and R.W. Pastor. Comparison of Langevin and molecular dynamics simulations. equilibrium and dynamics of ethylene glycol in water. *J. Chem. Soc., Faraday Trans*, 88:1747–1754, 1992.

[232] T. Wu. A study of convergence on the Newton-homotopy continuation method. *Applied Mathematics and Computation*, 168(2):1169–1174, 2005.

[233] T. Yamamoto. Historical developments in convergence analysis for Newton's and Newton-like methods. *The Journal of Computational and Applied Mathematics*, 124(1-2):1–23, 1980.

[234] H. Ye, W. Xian, and Y. Li. Machine learning of coarse-grained models for organic molecules and polymers: Progress, opportunities, and challenges. *ACS Omega*, 6(3):1758–1772, 2021.

[235] N. Yoshii and S. Okazaki. Molecular dynamics study of structure of clusters in supercritical Lennard–Jones fluid. *Fluid Phase Equilibria*, 144(1–2):225–232, 1998.

[236] H. Yukawa. On the interaction of elementary particles. i. *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series*, 17:48–57, 1935.

[237] Z. Zhang and R. Millan. Entropy production of non-reciprocal interactions. *Physical Review Research*, 5, 2023.

[238] R. Zwanzig and R.D. Mountain. High-frequency elastic moduli of simple fluids. *The Journal of Chemical Physics*, 43(12):4464–4471, 1965.