7-2020

A Study of the Thermodynamics of Small Systems and Phase Transition in Bulk Square Well-Hard Disk Binary Mixture

Gulce Kalyoncu
University of Arkansas, Fayetteville

A Study of the Thermodynamics of Small Systems and Phase Transition in Bulk Square Well-Hard Disk Binary Mixture

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Chemical Engineering

by

Gulce Kalyoncu
Ege University
Bachelor of Science in Chemical Engineering, 2009
Izmir Institute of Technology
Master of Science in Energy Engineering, 2014

July 2020
University of Arkansas

This thesis is approved for recommendation to the Graduate Council

_____
David M. Ford, Ph. D.
Thesis Director

_____
Gregory J. Thoma, Ph. D.
Committee Member

_____
Mahmoud Moradi, Ph. D.
Committee Member

ABSTRACT

Under the umbrella of statistical mechanics and particle-based simulations, two distinct problems have been discussed in this study. The first part included systems of finite clusters of three and 13 particles, where the particles are interacting via Lennard Jones potential. A machine learning technique, Diffusion Maps (DMap), has been employed to the large datasets of thermodynamically small systems from Monte Carlo simulations in order to identify the structural and energetic changes in these systems. DMap suggests at most three dimensions are required to describe and identify the systems with 9 (N = 3) and 39 (N = 13) dimensions. At the end of the study, a model has been proposed to functionalize the potential energy in terms of geometric variables that are identified with a heuristic screening. Investigation of the thermodynamics of bulk systems was another major focus of this thesis. The phase diagrams of the pure square-well solids and binary mixture of square-well and hard-disk particles, under the assumption of a pseudo-single-component model, have been constructed, and the phase equilibria behaviors were discussed. The datasets were also created in Monte Carlo simulations. The results showed isostructural solid-solid phase transition, which was previously identified that the pure square-well system with a very short range of attraction undergoes, also occurs in the presence of additional hard-disk components, namely for the binary mixture of square-well and hard-disk systems.

ACKNOWLEDGMENTS

In this section, I would like to acknowledge several people that have had influences on me whether by contributing to the intellectual property or showing guidance that made a significant impact on the quality of my thesis for the sake of improvement and progress.

Firstly, I would like to express my sincere gratitude to my supervisor Dr. David M. Ford. From the very first day that I have started working in his research group, the critical thinking, kind attention to detail, and the encouragement he has shown me to think outside the frame made me push myself even further to succeed. Moreover, thanks to his superb supervision, kindness, and patience, I have come a long way.

I would also like to thank my committee members for their navigation to enrich the context of my study. Firstly, Dr. Mahmoud Moradi was appreciated for the valuable conversations specifically about statistical thermodynamics that helped me grasp the theory more clearly and successfully project it onto my research. And secondly, Dr. Gregory J. Thoma was thanked for his sincere patience and availability in trying his absolute hardest to answer all of my questions about modeling and simulation. Also, I would like to appreciate Dr. Peter A. Monson's generosity for providing the Monte Carlo simulation code which was used in Chapter 3 after minor modifications.

Finally, I would like to thank my fellow graduate student colleague Aditya Dendukuri for his outstanding contributions in Chapter 2. The effort that he has put in generating the results for diffusion maps with IsoRank and Eigenspectra distance, which was an accurate distance metric brought to table by himself, have significantly heightened the quality of the study. Benjamin Drewry was also thanked as an undergraduate colleague for his contributions on the way.

DEDICATION

I would like to dedicate this thesis to my precious lifelong learning partner, my husband Efecan Pakkaner, who has always been incredibly supportive during the preparation of this thesis and devoted himself to encouraging me to persist my motivation. I will always feel privileged to share this path with him.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In this thesis, the behaviors of two different systems have been investigated from a thermodynamics and statistical mechanics point of view. In both studies, numerical investigations have been employed to the different characteristics of data sets that are produced via Monte Carlo simulations.

The first study was an investigation of the variables of thermodynamically small systems by using machine learning (ML) techniques. The focus was on atomic clusters having the number of particles equal to three and 13. Many researchers have focused on small clusters throughout the years due to the characteristics they possess that classical thermodynamics often fail to describe accurately. These mentioned characteristics differ significantly from the fundamental thermodynamic relations, such as having a range of coexistence temperature at the point of phase separation. In light of these and similar discrepancies between finite systems and bulk counterparts, the well-known thermodynamics conventionally applied to the bulk systems needs to be revisited, and new variables that are physically meaningful and applicable to the small systems should be identified in order to define the thermodynamic state of the system. For that particular purpose, small clusters have been simulated via Monte Carlo simulations with a constant number of particles, volume, and temperature (NVT). The output data contains the particle positions and their corresponding potential energy values. By using Diffusion Maps (DMap), which is one of the ML techniques that has been studied extensively with atomic or molecular clusters, output trajectories are embedded in the reduced subspaces. The possible relations of the subspaces with the order parameters that describe the structural and energetic changes in configurations are investigated. An extensive investigation has been conducted, and a numerical model of potential energy as a

function of the ‘best’ structural order parameters have been prepared. All details about the mentioned study are given in Chapter 2.

Chapter 3 of this thesis is dedicated to the investigation of the phase transition of the binary mixture of hard-disk and square-well particles. In contrast to the previously studied finite systems mentioned above, the thermodynamics of the bulk system has been the focal point of this chapter. For that purpose, a pure square-well system and 75% square-well (SW) 25% hard-disk (HD) binary mixture has been simulated by using Monte Carlo simulations. An extensive literature survey showed that the phase transition characteristics of pure square-well and pure hard-disk systems and their individual binary mixtures are widely studied and well documented. However, the best to our knowledge, the phase transitions in the binary mixture of square-well and the hard-disk system remains unstudied. Notably, it was shown that pure square well particles could exhibit an isostructural solid-solid transition for a very short range of attraction (Bolhuis, Hagen & Frenkel, 2004). This type of phase transition has an interesting feature as the configurational symmetry is preserved at different values of the lattice constant. It also resembles the first order liquid-solid phase transition by terminating at a critical point. Therefore, the main motivation behind the study was to see the effect of the additional hard-disk component on the square well solids on the phase transition behavior. In order to build the phase diagram of a binary mixture, free energies are calculated based on the thermodynamic integration methods that are given for the pure square-well systems (Bolhuis, Hagen & Frenkel, 2004). The rest of the study is explained in Chapter 3 in detail.

The outline of the thesis is built as follows. In Chapter 2, the ML technique to identify the variables of thermodynamically small systems has been explained and discussed with the results. Chapter 3 focuses on the phase transition of bulk hard-disk and square-well binary mixtures. The conclusions for each study are given at the end of their related sections.

CHAPTER 2

MACHINE LEARNING TO IDENTIFY VARIABLES IN THERMODYNAMICALLY SMALL SYSTEMS

2.1 Introduction

Over the past two decades, there has been a vast amount of research in the field of nanoscience, where the main focus is systems having dimensions on the order of the nanometers (Love et al. 2005). This growing interest in nanotechnology (Hill 2001; R. Stephen Berry 2002; Sehgal & Maroudas 2015; Marks & Peng 2016; Mohazzabi & Mansoori 2018; Grigoryan & Springborg 2019) emphasizes the necessity to fundamentally comprehend the characteristics of small clusters of particles such as atoms or colloids and shifts the focus of the related research from macroscopic to small systems.

As the interest in nanoscale applications rapidly increases in various scientific areas, thermodynamically small systems with a number of interacting particles N in the range of 1-1000 have been the focal point of numerous studies (Qian 2012). This interest calls attention to the discrepancies between the thermodynamics of small (finite) and infinite systems. As powerful as classical thermodynamics is, its ability to explain equilibrium behavior in small systems is inadequate, due to the lack of insights it provides for intermolecular interactions and energy fluctuations (Tester & Modell 1997; Baletto & Ferrando 2005). Thermodynamics of the small systems has been first discussed by Hill (1963), addressing the relation between bulk and small system thermodynamics, with the consideration of finite-size effects. A major conclusion of this new "Nanothermodynamics" is that various intensive properties are affected by the size of the system, where no similar effects of size were observed for macroscopic systems (Hill,1963;1964).

In order to re-cast the macro scale thermodynamic relations into a more generalized form that could be applicable for small systems as well, some modifications should be employed. As an example, differential changes in the Helmholtz free energy *A* for bulk systems may be expressed as,

$$dA = -SdT - pdV \tag{2.1}$$

where *S* is the entropy, *T* is temperature, *V* is volume, and *p* is the pressure. On the other hand, this equation needs an alteration if it is desired to be valid for the small systems such as liquid droplets, or colloidal particles. If the droplet or particle is still 'large enough', the addition of a new term in Equation 2.2, which shows the effect of surface forces, should suffice to describe the thermodynamics.

$$dA = -SdT - pdV + \gamma d\Lambda \tag{2.2}$$

Here, γ and Λ represent surface tension and the surface area, respectively. At this point, the major challenge is that utilizing the surface area or the volume as a variable for very small systems is not possible; in other words, the definitions themselves can be obscure. This situation creates a demand for new variables to be implemented to describe the thermodynamic relations for the small systems more accurately. Considering the characteristics of the systems and keeping in mind that the classical thermodynamics cannot prevail to come up with an explanation, statistical mechanics-based approaches that operate on particle-level simulations are required (Pan & Chandler, 2004). The output of the particle level simulations for *N* discrete particles is a set of particle coordinates ***X***, which is in a 3*N* dimensional space for 3-dimensional physical models. For each $\mathbf{r}_i$ that describes the position of the particle *i* in the Cartesian coordinates, ***X*** is shown in Eq. 2.3.

$$\boldsymbol{X}: \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N\} \in \mathbb{R}^{3N} \tag{2.3}$$

From the statistical mechanics point of view, the Helmholtz free energy is defined as follows,

$$A(N,V,T) = -k_B T ln Q(N,V,T) \tag{2.4}$$

where k_B is the Boltzmann constant. Q (N, V, T) is the partition function is defined as follows,

$$Q(N,V,T) = \sum_j \exp\left(-U_j(N,V)/kT\right) \tag{2.5}$$

where $U_j(N,V)$ indicates the potential energy of the microstate j, where the points in the phase space describes particle positions. (Hannon & Tromp, 2003).

As briefly shown in Equations 2.4 and 2.5, statistical mechanics could provide a straightforward route for the computation of Helmholtz free energy. However, for each different N and type of the particles, this calculation should be repeated, which yields a high computational cost. For an alternative analysis and understanding of the system, the dimensionality could be reduced into a form that is represented with a thermodynamic state equation such as Eq.2.2, by the aid of several variables that possess the collective properties of $\boldsymbol{X}$. Dimensionality reduction of these complex systems can be done via techniques that rely on disciplines including machine learning (ML) and statistics (Coifman et al., 2005; Das et al., 2006; Stamati et al., 2010; Ferguson et al., 2010, 2011).

In the current study, the output data sets of the canonical ensemble Monte Carlo simulation are processed with an unsupervised type of dimensionality reduction technique, Diffusion Maps (DMap) (Coifman et al. 2005),to identify the subspaces that reveal the underlying system characteristics. This method has been used in previous studies in order to recognize the subspaces in numerous kinds of finite systems (R. R. Coifman et al. 2008; A. L. Ferguson et al. 2010; Andrew et al. 2011; Rohrdanz et al. 2011; Xue et al. 2013; Nedialkova et al. 2014; Mansbach & Ferguson 2015; Dsilva et al. 2015; Reinhart et al. 2017).

This study can be divided into two main parts. First, the generation of the output trajectories of N particles, which includes the positions and potential energy data of the particles, occurs with the canonical ensemble MC simulation. Second, processing of the dataset to represent the data with the reduced coordinates has been executed, and the concluding remarks about order parameters have been stated. The main focus of this work is to study two different small systems with three and 13 particles interacting via Lennard Jones (LJ) potential. In the three-particle system (LJ3), the number of particles is selected considering the existing studies that have been performed for Argon clusters (Kaelberer & Etters 1976; Beck et al. 1988; Hinde, Berry & Wales 1992) Beck, Leitner, & Berry, 1988). Although a three-particle system seems extremely simple for a molecular simulation study, it is complex enough to demonstrate different structural behaviors to investigate the fundamental thermodynamic characteristics of small systems. The latter studied model system (LJ13), on the other hand, contains 13 interacting particles, which is known as one of the "magic numbers" based on the distinctive stability characteristic with respect to melting. Moreover, 13-particle clusters have broad phase coexistence ranges between solid-like and fluid-like structures, and, in comparison to the similarly small sized counterparts, they need higher temperatures to show significant fluctuations in distances between the particles (Beck et al. 1988; Beck, Berry & Becka, 1988). With the insight that working with both three and 13-particle systems provides, more accurate predictions of a scale-up to a system that has a larger set of particles could be made.

The main objective of this study is to discover the subspaces in $\boldsymbol{X}$ that collect essential information about the physical characteristics of the small systems LJ3 and LJ13 by applying DMap. After conducting DMap analyses for the two different systems, a set of collective variables that will formulate the thermodynamics are examined heuristically and presented.

2.2 Literature Review

In this subsection, studies that use the machine learning techniques to identify lower-dimensional subspaces that describe interesting structural transitions that are observed in very small clusters mostly having three and 13 particles are summarized. Additionally, a brief review of the studies that have performed DMap analyses on the molecular simulation trajectories is presented.

2.2.1 Thermodynamically Small Systems

As nanoscience and nanotechnology continue to advance in various application areas, including photonics, energy conversion, biological molecules, and condensed matter systems (Wang et al. 2017) methods of fabrication of these applications yield to even smaller systems of interest (De Miguel & Rubí 2017). These so-called "small systems," involving particle numbers of ten to a thousand, act as a unifying path between individual molecules and bulky macro systems (Jortner & Rao 2002). However, as we deviate from macro-scales to nanometric scales, classical thermodynamics becomes insufficient to explain the behavior of the systems as chemical and physical properties of these materials as well as with idiosyncratic structural characteristics substantially different from the bulk material characteristics (Jortner & Rao 2002). Therefore, understanding how a small system behaves thermodynamically is crucial to exploit the progress of the assembly of those nano-systems to produce collective structures, including arrays, assemblies, and super-lattices.

Studies about the structural changes have a significant role in comprehending the features of the non-macroscopic systems, and the rich variation of the structural changes that occur in the course of phase transition. Hence, the reviewed studies provided reference points about the temperature values used in this study, although this study is not mainly focused on the phase

transitions of the small system. The variations in different structures that the system has, enables us to collect the dataset, which has rich characteristics to be further detected by machine learning techniques.

The uniqueness of small systems becomes obvious during phase transition; such that, the discrete temperature where the phase transition occurs at the constant pressure in bulk systems becomes a range of temperatures for the significantly small counterparts, and creates ambiguity between first and second-order transitions (Borrmann, Mülken, & Harting 2000; R.; Berry, 2002; Reguera et al. 2003; Palomares-Baez, Panizon, & Ferrando 2017; Philippe 2017). Transitions between solid-like and liquid-like structures in the small clusters having a number of particles as low as three have been shown by using Monte Carlo simulations (Kaelberer & Etters 1976). A study by Beck and Berry investigated the behavior of small clusters having a varying number of particles from seven to 33 in the constant energy ensemble during the phase transition. Molecular dynamics simulations were used, particularly focusing on the melting-freezing transition, by changing the number of particles. They also pointed out three magic number clusters (N = 7, 9, and 13) with the results of the potential minima on the surface have stable isomers when they are neighboring with the excited isomers. Also, during the melting stage, frequently observed structures are the ones that possess the pentagonal symmetry (Beck & Berry, 1988). The same structure was observed in the melting stage later by Hoare and Pal (2006). For 13 particle Lennard Jones clusters (Kaelberer & Etters 1976; Quirke & Sheng 1984; Jellinek et al. 1986; Honeycutt & Andersen 1987; Davis, Jellinek & Berry 1987; Berry et al. 1988; Beck et al. 1988; Berry & Smirnov 2009; Calvo, Parneix & Basire 2009) melting has been observed as the configuration morphed from icosahedral to a non-crystalline structure(Quirke & Sheng 1984). Intriguing structural transition characteristics have also been analyzed with the heat capacity curves of the

related systems, and the phase transition for such small systems have been unexpectedly found in comparison to the well-known first order phase transitions (R Stephen Berry, Jellinek & Natanson 1984; Quirke & Sheng, 1984; Davis, Jellinek & Stephen Berry, 1987). An exhaustive review of the phase transition of different size clusters can be found elsewhere (Berry & Smirnov, 2009).

2.2.2 Diffusion Maps

Diffusion Maps (DMap) (Coifman et al., 2005) is a non-linear machine learning technique that has been studied extensively in recent years and has been proven to be effective as a non-linear dimensionality reduction technique (Koronaki, Nikas & Boudouvis, 2020). Moreover, its versatility has been confirmed as its successful applications has been previously published in various research areas including meteorology (Fernandez et al., 2015) data analysis and algorithm improvement in flight simulators (Jia et al., 2016), academic data mining and informetrics (Nieminen, Polonen & Sipola, 2013), genome and biomedical engineering (Xu et al., 2010; Galdino, Benevides & Tenorio, 2020) and life sciences (Wai et al., 2009) along with numerous other adaptations.

DMap has also gained significant interest in stochastic and multi-scale dynamic system interpretations, as the study by Nadler et al. focused on implementing DMap for the evaluation of predicting reaction coordinates in a particle-based reaction system (Nadler et al., 2006). It was concluded that diffusion maps naturally are impactful in analyzing the empirical datasets generated from an unknown probability distribution. Diffusion maps were also found to be a powerful tool for analyzing dynamic systems such as reaction and particle systems, with different scales of time, and characteristics of the dynamic systems can be successfully characterized in the long term by considering the top eigenfunctions. Another work based on a similar understanding was published by Koronaki and co-workers (Koronaki, Nikas & Boudouvis, 2020), where an effort was made to

combine both DMap and neural networks (NN) to construct a low-order model for a non-linear system. In the mentioned study, a tubular reactor with multiple steady-state cases was analyzed, and an accurate low-dimensional projection of the datasets both derived by simulations and experiments was constructed. It was concluded that the combination of DMap and NN was successful in representing the non-linear data in a lower-dimensional manifold. DMap has also been employed to identify the number of variables to create dynamic models in colloidal systems (Bevan et al., 2015; Yang et al., 2016). To model the crystallization dynamics of colloidal clusters, DMap has been used to investigate the required number of variables in another study. After a heuristic analysis based on the author's previous studies (Beltran-Villegas & Bevan, 2011; Beltran-Villegas et al., 2011), the order parameters were identified as; the radius of gyration of the cluster, and the crystalline parameter which briefly is the average number of neighboring particles in expected lattice locations that each particle possesses (Beltran-Villegas et al., 2012). The same order parameters have been found to be proper to identify the solid-solid transition at low temperatures, and the fluid-solid transition at high temperatures after the number of variables has been detected by DMap for an LJ38 cluster (Sehgal, Maroudas & Ford, 2014). Hub and de Groot (2009) reported a method, which is called as 'functional mode analysis' (FMA) to interpret the relations between collective motions and a particular protein function. The protein function is basically identified with its geometric, electrostatic, or chemical properties. The relation between functional quantity and collective motion was proposed to build by using two ways, such as using the motion that exhibits the highest correlation to functional quantity and the motion which performs large variations in the functional quantity.

DMap was also utilized to improve the already-established techniques, as the study by Lancewicki and Aladjem (2014) focused on enhancing the locally multi-dimensional scaling

(LMDS) by performing clustering in DMap to construct so-called "neighborhoods" in the high-dimensional dataset. Different than the original LMDS method chosen, the lower-dimensional DMap data revealed that the system could be interpreted with a fewer number of parameters to be selected by the user, thus yielding an easier interpretation of the mapping. As a promising approach to use DMap in bioprocesses, the folding mechanism of Trp-cage, a helical mini-protein, was investigated using unbiased MD simulations and DMap in work by Kim et al. (2015). The MD dataset was analyzed and via DMap, by projecting the MD dataset onto a lower-dimensional space, and the folding-unfolding pathways of the select protein were characterized. Although minor discrepancies were observed between computational and experimental findings, it was concluded that DMap was a robust tool to extract the underlying characteristics of the folding sequences in a lower-dimensional space, without the need of any other physical order parameters. Lastly, in an exhaustive work by Ferguson and co-workers, the approach to use DMap as a dimensionality reduction technique was reviewed, and its applications in the branch of molecular simulations were discussed (Ferguson et al., 2011). In summary, the study implies that DMap is a powerful tool to extract low-dimensional embeddings from high-dimensional simulation trajectories, and crucial dynamical modes can be discovered. The work also mentions several application examples from chemical and biochemical processes.

2.3 System Description and Computational Method

2.3.1 System Description

The particles are modeled with the pair-wise Lennard-Jones potential, u_{ij}, as shown in Eq. 2.6. These interactions are categorized as attractive and repulsive forces, which are presented by the terms $(\sigma/r_{ij})^6$ and $(\sigma/r_{ij})^{12}$, respectively.

$$u_{ij} = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{2.6}$$

where $r_{ij} = \left|\mathbf{r}_i - \mathbf{r}_j\right|$ defines the distance between particles *i* and *j*, and $\mathbf{r}_i$ is the Cartesian coordinates of particle *i.* The terms ε and σ are the Lennard Jones energy and size parameters, which vary based on the atom or molecule. Figure 2.1 represents the LJ potential as a function of the distance between particles, where for the short pair-wise distances, particles have repulsive potential caused by the overlap between the electron clouds of the atoms presented (Parker, 2001). If the distance between particles is large, atoms have the attractive van der Waals forces (Vlugt 2009). Specifically, LJ potential has its minimum when $r_{ij} = 2^{1/6}\sigma$. LJ potential has been used in a vast number of studies as seen in Section 2.2.1, and global minimum energies of clusters with a different numbers of particles interacting via LJ potential are well documented (Hoare & Pal, 1971; Wales & Berry, 1990).

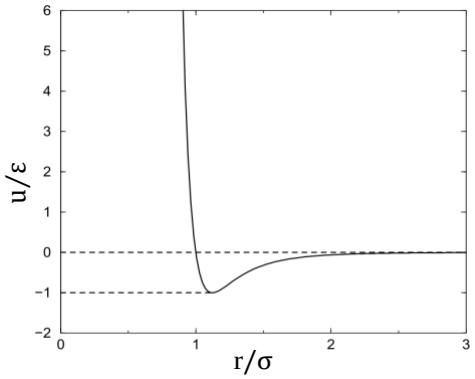

Figure 2.1. LJ potential as a function of displacement (source: Vlugt et al., 2009)

In the current study, interacting particles in a three-dimensional box, which is 10σ on a side, are simulated in the canonical ensemble, where the number of particles *N*, the volume *V*, and the temperature *T* of the system are kept constant throughout the simulation. During the simulation, the units of length and temperature are defined as σ, and k_B/ε, respectively. The total potential

energy of the particles U is the summation of the u_{ij} between particles defines the potential energy of the configuration and can be computed as follows.

$$U = \sum_{i=1}^{N-1} \sum_{j=j+1}^{N} u_{ij} \tag{2.7}$$

From now on, all variables will be reported in the non-dimensional form, including the non-dimensional temperature, which is defined as $T^* = k_B T/\varepsilon$, and dimensionless potential energy on a per-particle basis, that is given as $u^* = U/N\varepsilon$. For the LJ3 system, T^* is selected as 0.18. On the other hand, two different temperature values as 0.28 and 0.4 are examined for the LJ13. The selection criteria of the temperature are explained in Section 2.3.3.2. In the LJ3 system, initial configuration is chosen as triangular structure of minimum energy with $u^* = -3$, while an initial configuration, which gives the global minima for the icosahedron structure with $u^* = -3.4097$ is used for the LJ13 system (The Cambridge Cluster Database (CCD), n.d.). The total length of the simulation is set to 10^7 simulation steps, and with the sampling, 5000 configurations are collected for further data analysis.

2.3.2 Structure Dependent Variables - Identification of the Structural Variations

During the simulation, there are changes in the size and shape of the configurations in the clusters. This study aims to describe the possible thermodynamic variables of the small systems. Therefore, some physical variables that can be used to characterize the structural variations of small clusters are described below.

The radius of gyration r_g is an essential parameter that has been significantly used to identify the structural changes that may occur in the clusters. This parameter explains to what extent the cluster is dissociated or condensed. Each clusters' intrinsic size could be found by computing their r_g values (Lerbret et al., 2005). As a well-known fact from the literature, the

previous studies have shown that the r_g is a good parameter to deduce useful information about the configurational changes in small systems (Beltran-Villegas & Bevan, 2011; Beltran-Villegas et al., 2011; Beltran-Villegas et al., 2012). r_g can be calculated with the following formula:

$$r_g = \frac{1}{N}\sqrt{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left\|\mathbf{r}_i - \mathbf{r}_j\right\|^2} \tag{2.8}$$

Another key variable for capturing the structural changes is the number of bonds (n_b) between the particles. The calculation of the n_b can be found in the Eq. 2.9.

$$n_b = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} f_c(r_{ij}) \tag{2.9}$$

where $f_c(r_{ij})$ is zero if the Euclidean distance between the particles is larger than the cutoff distance, and one, if the pair-wise distance is smaller than the cutoff distance. For both systems, the cutoff distance was selected as 1.2σ.

The final parameter that was included in this study is the connectivity C, which can be described as the average of the number of bonds of each particle is defined in Eq. 2.10.

$$C = \frac{n_b}{N} \tag{2.10}$$

2.3.3 Computational Methods

2.3.3.1 Monte Carlo Simulations

Particles are simulated by using the Monte Carlo (MC) simulations. MC is a practical way of simulation for the cases where the dynamics of the system are not considered, but only the static (equilibrium) properties are the matter of interest. In the current study, the main motive behind using MC simulations is to sample different structures that the small systems with three and 13 particles hold in an ensemble. Thus, it is not necessary to take the time into account.

The trajectory output dataset was generated by the Metropolis algorithm (Metropolis et al., 1954). This algorithm is built upon the acceptance of the particles moving based on a probability

value of $e^{-\beta\Delta u}$, where β and ΔU are the inverse temperature and the potential energy difference between two consecutive steps, respectively. One important criterion in the use of the Metropolis algorithm is the selection of the maximum displacement $\delta r_{max,}$ which develops the size of the trial move and is responsible for the acceptance of the moves by checking the convergence of the Markov chain. Extra attention should be given to select δr_{max}, since for the significantly small values of δr_{max} almost all moves are accepted, and conversely, for very large values of δr_{max}, a high percentage of the moves are rejected. In this study, δr_{max} is identified as 0.15σ initially, and this value is adjusted throughout the simulation to maintain the 50% acceptance rate of the trial moves. The initial value of δr_{max} should not be set as more than the half-box size and the selected δr_{max} value for this study is one of the typically used values in the literature (Vlugt,2009; Dijkstra, 2014).

In the simulation of the mentioned systems, periodic boundary conditions (PBCs) are used to maintain the periodicity with the assumption of the unit cell is replicated to model an infinite space. Minimum image convention approach was used in the calculation of the pair-wise separation distances between the particles to apply PBCs. Since there are no solid boundaries between the central unit cell and the replicas, a particle can travel from the central unit cell to one of the neighboring boxes in any direction. However; as all the atomic movement is known to be replicated by the neighboring unit cells simultaneously, another particle from the other neighboring replica having the same potential energy comes across the boundaries to the main unit cell, resulting in conservation in the total number of particles (Allen & Tildesley, 2017).

2.3.3.2 Data Generation

To generate the dataset, a canonical ensemble Monte Carlo simulation code from the literature (Allen & Tildesley, 2017) was used by implementing minor modifications. The output potential energy values are validated against literature data where available. For the LJ13 system, corresponding potential energy values were already reported by Kaelberer and Etters (1977), and the current study simulation results are found to be in good agreement (Fig. 2.2). As, the potential energy increases gradually, the slope of the curve undergoes some discrepancies in the range of T^* = 0.25 and 0.30 values; also, at T^* = 0.4, which might suggest the phase transitions between solid-like and liquid-like structures, and a liquid-to-vapor transition could be occurring in the said T^* values. In light of these, the temperature values were selected as 0.28 and 0.4 for the simulation. The literature data stopped at T*=0.28. A more detailed investigation about the possible transitions at the mention temperatures is done with the help of heat capacity computations and will be given further on.

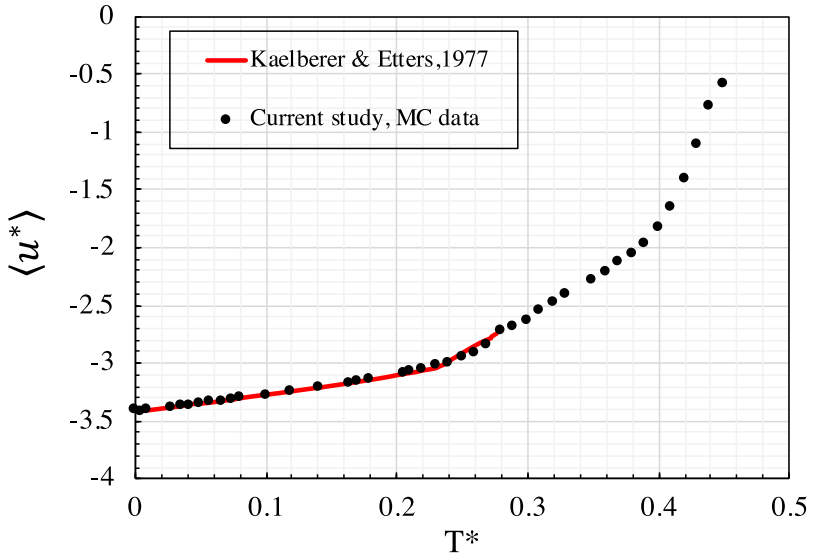

Figure 2.2. Validated results of dimensionless potential energy values with reference studies.

To investigate the statistical significance of the validation, an unpaired t-test was applied to the simulation data and data acquired from the literature for validation (GraphPad, San Diego, CA, USA). As a result, the statistical difference between two datasets was found to be not significant, with an overall *p*-value of 0.8723 (N = 25).

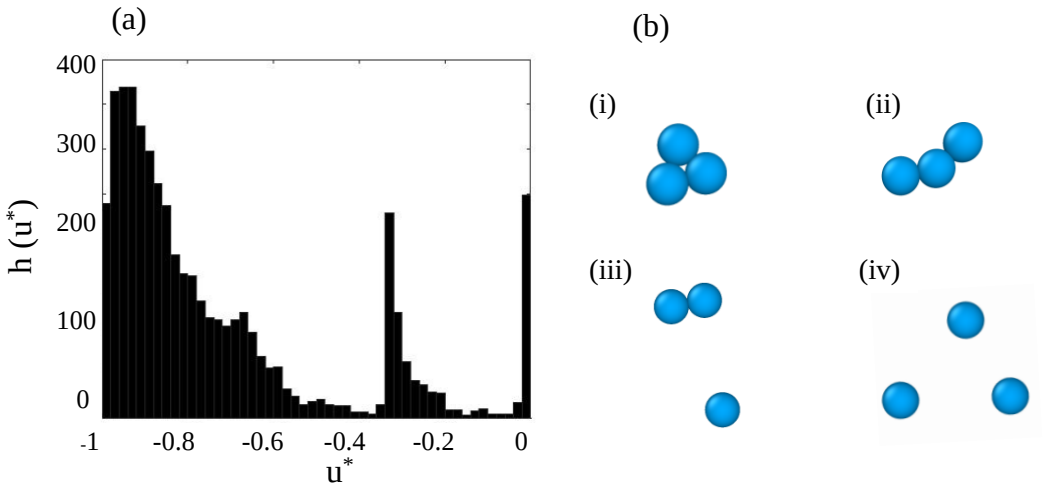

Figure 2.3. LJ3 system, T^* = 0.18 (a) Histogram of dimensionless potential energy u^* (b) different structures observed at T^* = 0.18 (i) tightly clustered, (ii) linear, (iii) partially dissociated, and (iv) completely dissociated.

The value of the temperature is crucial since it directly affects the nature of the data. In order to identify the number of structures sampled by the system, an extensive analysis of the system's potential energy distribution at varying temperature values has been performed. The range of the potential energy values that the systems have, and the heat capacity curve can be used as the indicators of the different phases. Figure 2.3(a) shows that the LJ3 system has different u^* values varying from -1 to 0 when T^* is equal to 0.18. This was stated by Ford et al. (2020) and quoted as follows. "It is noteworthy to point out that seeing positive values of u^* is possible when one or more of the pair-wise distances are less than σ; however, they are rather uncommon at this relatively low-temperature. In the data set, the maximum value observed was $u* = 0.00013$, and very few configurations had $u* > 0$ so that no counts in that region are

visible on the scale of the histogram" (Ford et al., 2020). In Figure 2.3(b), four different structures varying from tightly clustered to completely dissociated that the LJ3 system has experienced during the simulation can be seen. Consequently, T^* was selected as 0.18, because all of the structural motifs and potential energies were observed in a single molecular simulation trajectory.

An extensive study to identify temperature values that are rich in terms of the different structures has also been done for the LJ13 system. Similar to the 3-particle system, potential energy histograms are used as helpful tools to decide on the value of temperature. Figure 2.4 represent the potential energy distributions for LJ13 system at $T^* = 0.28$ and $T^* = 0.4$. The configurational snapshots are created with Open Visualization Tool (OVITO) (Stukowski,2010).

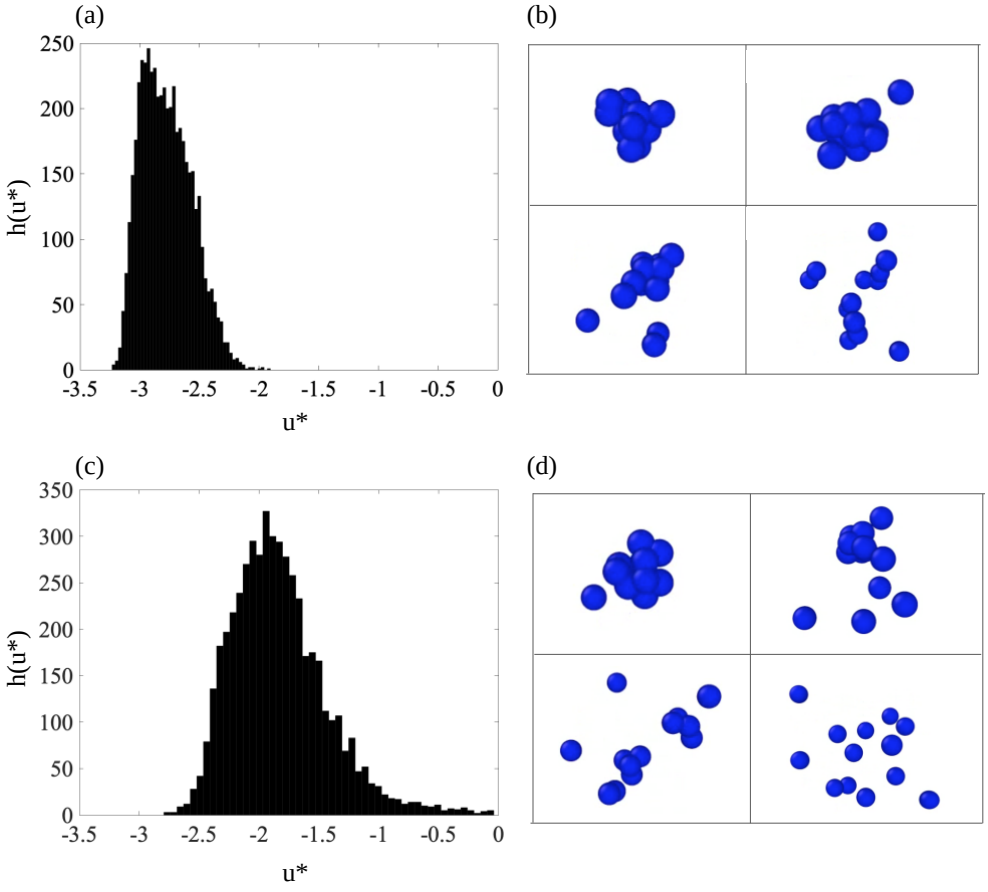

Figure 2.4. LJ13 system. Histograms of dimensionless potential energy u^* at (a) $T^* = 0.28$ and (c) $T^* = 0.4$. Some select structures that have been observed for LJ13 system at (b) $T^* = 0.28$, (d) $T^* = 0.4$.

Besides the wide range of the potential energy, and the existing literature about the phase transition of Ar13 studies, heat capacity values have been calculated for these specific temperature values with the formula given in Eq. 2.11 (Davis, Jellinek, and Stephen Berry 1987).

$$C_v = \frac{\langle u^2 \rangle - \langle u \rangle^2}{kT} \tag{2.11}$$

The case when T^* equals 0.28 has already been reported as a transition temperature for Ar13 clusters in the literature, as a result of having a peak in the heat capacity curve (Davis, Jellinek, & Berry, 1987). The current study results also represent a peak in the heat capacity versus temperature plot, as it was shown in Figure 2.5. The first peak at $T^* = 0.28$ may be indicating that there could be a transition occurring between solid-like and liquid-like configurations. The next peak in the heat capacity curve is observed at the temperature around 0.4, where the particles exhibit more broken structures. Therefore, it can be classified as a liquid-like to the gaseous-like phase transition.

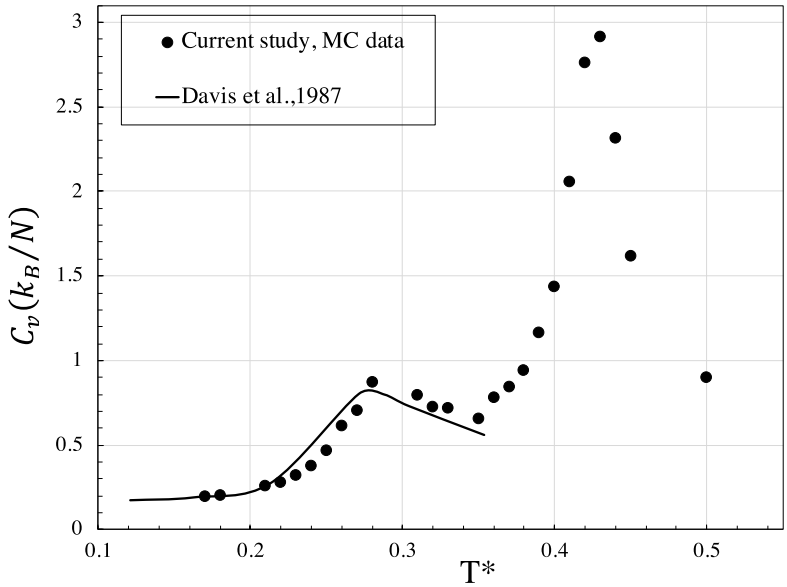

Figure 2.5. Heat capacity values as a function of temperature for LJ13 system.

2.3.3.3 Dimensionality Reduction

The output trajectories represent the spatial configurations of the *N* particles and their corresponding potential energies, which are obtained at the *3N* dimensional system. A total number of 5000 configurational 'snapshots' that were collected throughout each simulation run treated by ML techniques in order to investigate the characteristics of those "thermodynamically small systems" (Hill, 1963) to display that it is possible to identify the order parameters which then can be used to describe the potential energy. Based on the non-linear characteristics of the output trajectory, an unsupervised ML technique, DMap is applied to the data set. The method can be briefly explained as the re-organization of the data based on a diffusion metric by mapping the data from the original coordinates to the diffusion space (Coifman & Lafon, 2006).

The construction of DMap mostly relies on the principles of Markov's random walk and the diffusion theory. As the first step in the DMap algorithm, Gaussian kernel matrix K_{ij} should be assembled by considering pairwise distance between set of particles in configurations *i* and *j*.

$$K_{ij} = \exp\left(-\frac{d_{ij}^2}{2\delta^2}\right) \tag{2.12}$$

After constructing the kernel matrix, the kernel bandwidth (δ), which has the role of describing the local connectivity between data points, should be evaluated before proceeding to the rest of the algorithm. The selection of this parameter is important in terms of having a proper characterization of the data set in mapping onto the reduced space. Low values of kernel bandwidth represent larger distances between pairs of the data points, which is an indicator of loosely connected behavior and returns to small kernel sum. On the other hand, for the relatively higher values of δ, the scaled distance between data points decreases, and consequently, the data points become well-connected in a pair-wise sense. The linear region in the plot of $\log(\sum K)$ as a function of $\log(\delta)$ bridges those two extremes is the region of interest, as it captures the natural connectivity

of the data at intermediate scales, and the value of δ is selected from this region, as will be discussed later.

For each data point, the summation D_{ii} of the particular data point and all other points are found, and the Kernel matrix is normalized dividing by this summation.

$$D_{ii} = \sum_{j} K_{ij} \tag{2.13}$$

Markov matrix which describes the probability of the transition from one particular configuration to another, is constructed as:

$$M_{ij} = \frac{K_{i,j}}{D_{ii}} \tag{2.14}$$

The eigenvalues λ and right eigenvectors ψ shown in Equation 2.15 is calculated with the eigenvalue decomposition.

$$\mathbf{M}\boldsymbol{\psi} = \lambda\boldsymbol{\psi} \tag{2.15}$$

One of the crucial steps in the DMap algorithm is the selection of the type of distance calculation to obtain the kernel matrix given in Equation 2.12. The definition of the distance affects the ultimate results based on the inherent characteristics of the data, and the investigation of these effects is one of the major purposes of this study. In the current study, in order to examine the impact of distance type, four different types of distance calculations are used.

The first definition is the Hausdorff distance, which is a relatively general and straightforward metric for the data points in the Euclidean space. To calculate this distance, the Euclidean distance from each discrete particle in one configuration to all the particles in the other configuration is computed, and the shortest of these distances is determined. The greatest of these distances gives the Hausdorff distance between the two configurations. Mathematically, this can be represented as (R. R. Coifman et al. 2008),

$$d(S_1, S_2) = max\left\{\max_{1\le i\le N}\min_{1\le j\le N}\left\|S_1^i - S_2^i\right\|, \max_{1\le i\le N}\min_{1\le j\le N}\left\|S_2^j - S_1^i\right\|\right\} \quad (2.14)$$

where S_1^i and S_2^i are the positions of i_{th} and j_{th} particles in the configurations S_1 and S_2, respectively.

The second distance definition is constructed based on an adaption of Mayer *f*-bond from statistical mechanics (Mayer, 1942). Each configuration is assigned with a numerical value called F_m (Eq. 2.15), and the distance metric is constructed by calculating the absolute difference between configurational F_m values.

$$F_m = 1 - \exp(-u^*) \quad (2.15)$$

As F_m is an explicit function of u^*, and u^* does not precisely discriminate between structural motifs as shown in Figure 2.3(a), it is not expected this distance metric to solve the problem at hand but rather to serve as a kind of test case. For this metric, the DMap analysis should collapse the data to one dimension that correlates perfectly with u^*. The descriptions of IsoRank and Eigensprecta distance metrics were given previously in Ford et al. (2020), and I quote them here. "The third distance metric is based on IsoRank (Singh et al., 2008), which is an algorithm to detect the best alignment between two given plot structures. IsoRank was first used as a distance metric for DMap calculations for molecular clusters in work by Long and Ferguson (2014), to counter the lack of an invariant basis in the dataset when the particles are indistinguishable. This distance metric utilizes two adjacency matrices. The first matrix ($\mathbf{R} \in \mathbb{R}^{N\times N}$) holds the squared Euclidean distance between particles *i* and *j* ($\mathbf{R}_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|^2$) while the second matrix ($\mathbf{G} \in \{0, 1\}^{N\times N}$) is a binary connectivity matrix representing the existence of or lack thereof a bond between the two particles that can be defined based on a cutoff distance (used as 1.2σ). IsoRank is a metric that computes an alignment score between two configurations of unlabeled particles. The first step of the distance computation is to compute the scores for all possible permutations of labels and pick

the best alignment. This step is the main bottleneck of this technique, as one must get the alignment scores for every possible permutation before selecting the best alignment. For snapshots *i* and *j*, let $\mathbf{R}_{\mathrm{i}}^{*}, \mathbf{G}_{\mathrm{i}}^{*}, \mathbf{R}_{\mathrm{j}}^{*}, \mathbf{G}_{\mathrm{j}}^{*}$ be the best alignment detected by IsoRank. Then, the distance kernel for building the diffusion mapping is calculated by the sum of the absolute differences between the inter-particle distances as follows,

$$d_{ij} = \sum_{p=1}^{N} \sum_{q=p+1}^{N} \left| \mathbf{R}_{\mathrm{i}}^{*}(p,q)\mathbf{G}_{\mathrm{i}}^{*}(p,q) - \mathbf{R}_{\mathrm{j}}(p,q)\mathbf{G}_{\mathrm{j}}(p,q) \right| \tag{2.17}$$

The algorithm clips away particles that are far away by multiplying the distance with the binary bond value. This step ensures explicit separations of particles that are broken off a cluster. This also tends to generate a discrete subspace due to the sparse nature of **G**.

The fourth distance metric, Eigenspectra distance, borrows the idea of using an adjacency matrix representation to compare two clusters from IsoRank. Instead of using two adjacency matrices (one holding distances and one holding connection values), only one matrix $\mathbf{H} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ is defined that holds a value between 0 and 1 representing an implicit notion of the bond strength. This metric is computed by normalizing the pair-wise distances between 0 and 1. Let $\mathbf{H}_i$ and $\mathbf{H}_j$ be the bond-strength matrices for two configurations with indices *i* and *j* respectively and $\mathbf{\Psi}_i\mathbf{\Lambda}_i\mathbf{\Psi}_i^{\mathbf{t}}$ and $\mathbf{\Psi}_j\mathbf{\Lambda}_j\mathbf{\Psi}_j^{\mathbf{t}}$ be their respective eigen decomposition ($\mathbf{\Lambda}$ is a diagonal matrix holding the eigenvalues of **H**, with $\mathbf{\Psi}$ holding the respective eigenvectors). One important thing to note is that **H** is semi-definite, which means that the eigenvalues are always real numbers. The distance between configurations *i* and *j* is then computed as the absolute difference between the eigenvalue matrices as,

$$d_{ij} = Tr\left|\mathbf{\Lambda}_i - \mathbf{\Lambda}_j\right| \tag{2.18}$$

where *Tr* represents the trace operation. This is different from the approach used in the IsoRank-based distance as instead of aligning the graph, eigen decomposition is performed to compare the geometric properties of the graphs and counter the lack of invariance. The use of eigenvalues to compare the structures of two graphs is inspired by the established mathematics of networks (Straffin, 1980). The asymptotic complexity for eigen decomposition is $O(N_3)$, with N being the number of particles, as compared to IsoRank, which scales exponentially with *N* due to the particle-index permuting step" (Ford et al., 2020).

Computations for Hausdorff and Mayer f -bond distances were done using MATLAB (MathWorks, Natick, MA, USA), and computations for IsoRank and Eigenspectra distances were done using Python 3.0 with NumPy for the linear algebra calculations.

2.4 Results and Discussion

In the following section, the results are exhibited for both LJ3 and LJ13 systems at the corresponding temperatures. The datasets are represented in the reduced plane, and the variables which were found after the heuristic study for describing the thermodynamics of the systems are discussed.

2.4.1 Pre-processing of the Data

Considering that the system has 3*N* dimensionality, for three-particle and 13-particle LJ systems, dimensionalities are calculated as 9 and 39, respectively. Before employing DMap to the mentioned datasets, translational and rotational degrees of freedom of the systems should be removed in order to focus on internal structure solely. The purpose of removing these degrees of freedom was explained in the study by Ford et al. (2020): "Removing translational and rotational degrees of freedoms serves to minimize differences between configurations that are based purely on absolute location and orientation within the simulation box; thus, allowing the subsequent

analyses to identify the internal differences in the structure. Most of the analyses employed here, with the exception of DMap using the spectral distance and IsoRank metrics, would be sensitive to differences in absolute location and orientation, and these pre-processing operations also reduce that sensitivity" (Ford et al.,2020). The raw output data and pre-processed version of the same datasets are presented in Figure 2.6 both for LJ3 and LJ13 systems.

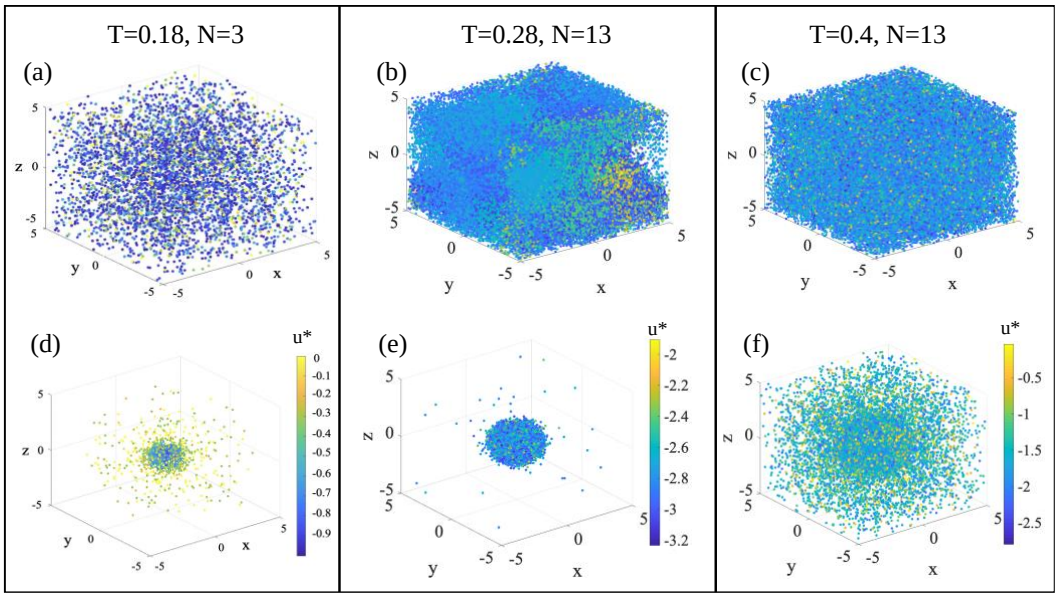

Figure 2.6 Representation of the output trajectory obtained for 5000 configurations. Unprocessed data for (a) N = 3, T* = 0.18, (b) N = 13, T* = 0.28, (c) N = 3, T* = 0.4. Configurations after mean-centering and aligning by their principal axes of rotation for (d) N = 3, T* = 0.18, (e) N = 13, T* = 0.28, (f) N=3, T* = 0.4. Data is colored based on the deviations in the potential energy.

2.4.2 Eigenvalue Spectrum Based on Different Distance Metrics and Dimensionality

After the pre-processing, the DMap algorithm was applied to the dataset with varying distance metrics. As the first step of the DMap method described in the previous section, a proper kernel bandwidth should be identified first. For that, the method which was proposed by Coifman et al. (2008) was used, and $\log(\Sigma K)$ vs. $\log(\delta)$ plots were drawn for both LJ3 and LJ13 systems. Kernel bandwidth values were selected from the linear regions of the sigmoidal shape curves, and corresponding values were exhibited in Figures 2.7 and 2.8 for LJ3 and LJ13 systems, respectively.

Kernel bandwidth values that are selected from the linear region of the sigmoidal shape curve for each case are indicated with the arrow in the sub-figures of in Figures 2.7 and 2.8. Some investigations have been completed before deciding which δ value from the linear region is best in terms of showing relations between possible order parameters. The preliminary studies showed that the results obtained by using the δ value from the edge of the linear region can show interesting relations with order parameters that have not been observed for their counterparts from the middle of the linear region.

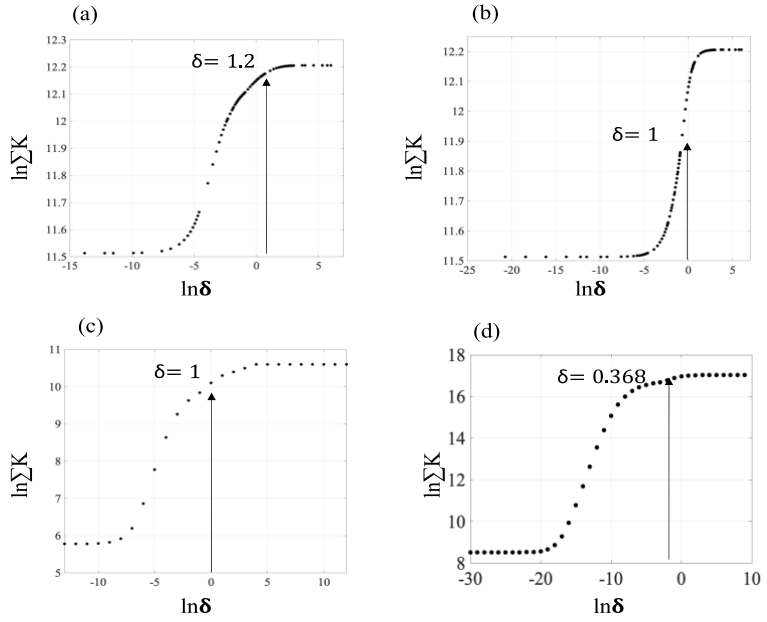

Figure 2.7. Identification of the kernel bandwidth with different distance metric types for the three-particle system. (a) Hausdorff Distance, (b) Mayer *f*-bond, (c) IsoRank Distance, (d) Spectral Distance.

The first seven eigenvalues calculated as a part of the solution for the right eigenvalue problem given in Equation 2.15 concerning different types of distance metrics are exhibited in Figures 2.9 and 2.10 for $N = 3$, and $N = 13$. By the aid of these plots, and considering the spectral gap between the data points, the total dimensionality of the system can be estimated for each particular system analyzed with the different distance metrics. The number of eigenvalues that yield to the most significant spectral gap between the current and successive eigenvalue presents

the number of dimensionalities. Based on this approach, the number of dimensionalities, that is required to comprehend the system characteristics fully and to completely analyze the system is suggested to be at most three for all types of distance metrics for both LJ3 and LJ13 systems.

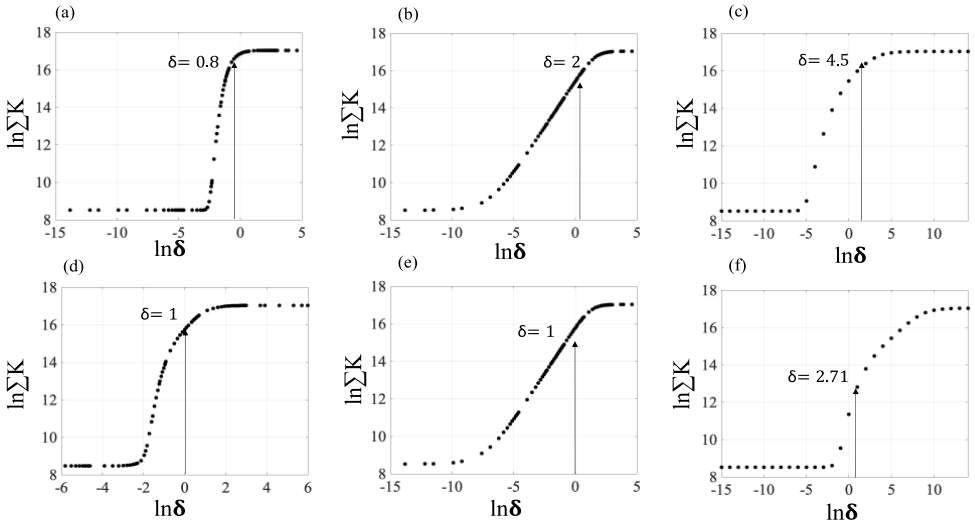

Figure 2.8. Identification of the kernel bandwidth with different distance metric types for the 13-particle system. $T^* = 0.28$ for the figures at the top, $T^* = 0.4$ for the figures at the bottom. (a) and (d) Hausdorff Distance, (b) and (e) Mayer *f*-bond, (c), and (f) Spectral Distance.

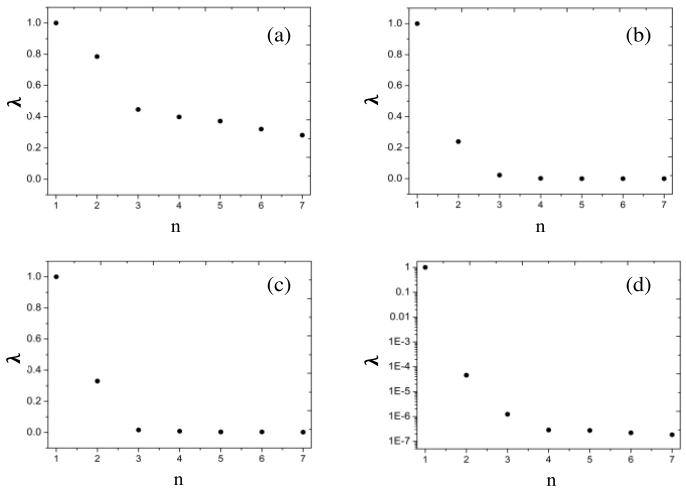

Figure 2.9. First n number of eigenvalues for different distance metric types when $N = 3$. (a) Hausdorff Distance (b) Mayer *f*-bond distance (c) IsoRank (d) Spectral Distance.

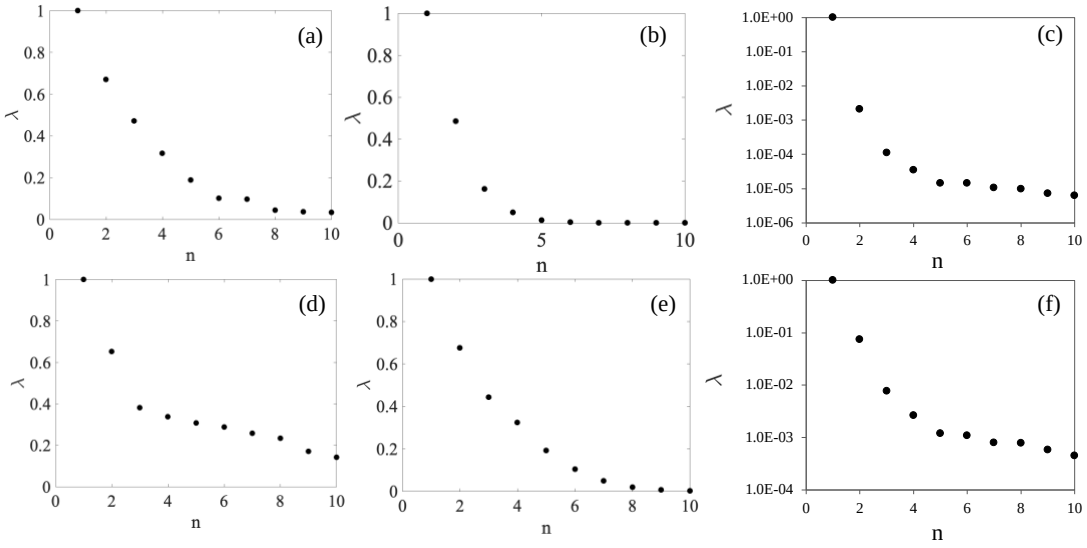

Figure 2.10. Eigenvalue spectrum for different distance metric types when $N = 13$. $T^* = 0.28$ at top, $T^* = 0.4$ at the bottom. (a) and (d) Hausdorff Distance, (b) and (e) Mayer *f*-bond distance, (c) and (f) Spectral Distance.

2.4.3 Embedding the Data in the Lower Space

In this section, the datasets for LJ3 and LJ13 particles are represented in the reduced subspaces, which are directly linked to the eigenvectors found at the end of the DMap analysis. The eigenvectors are examined in terms of the different distance metrics, and detailed results are represented. The analyses are performed by looking at the correlations of the eigenvectors with the dimensionless potential energy.

Figure 2.11 demonstrates the datasets mapped onto the reduced space, which are the final results of the DMap with the different diffusion metrics for $N = 3$. Although the dimensionality of the system has been concluded to be not more than three, the data was represented in three dimensions for the sake of a better visualization. For mapping the data onto the reduced space, the top non-trivial eigenvectors are used as the new coordinates.

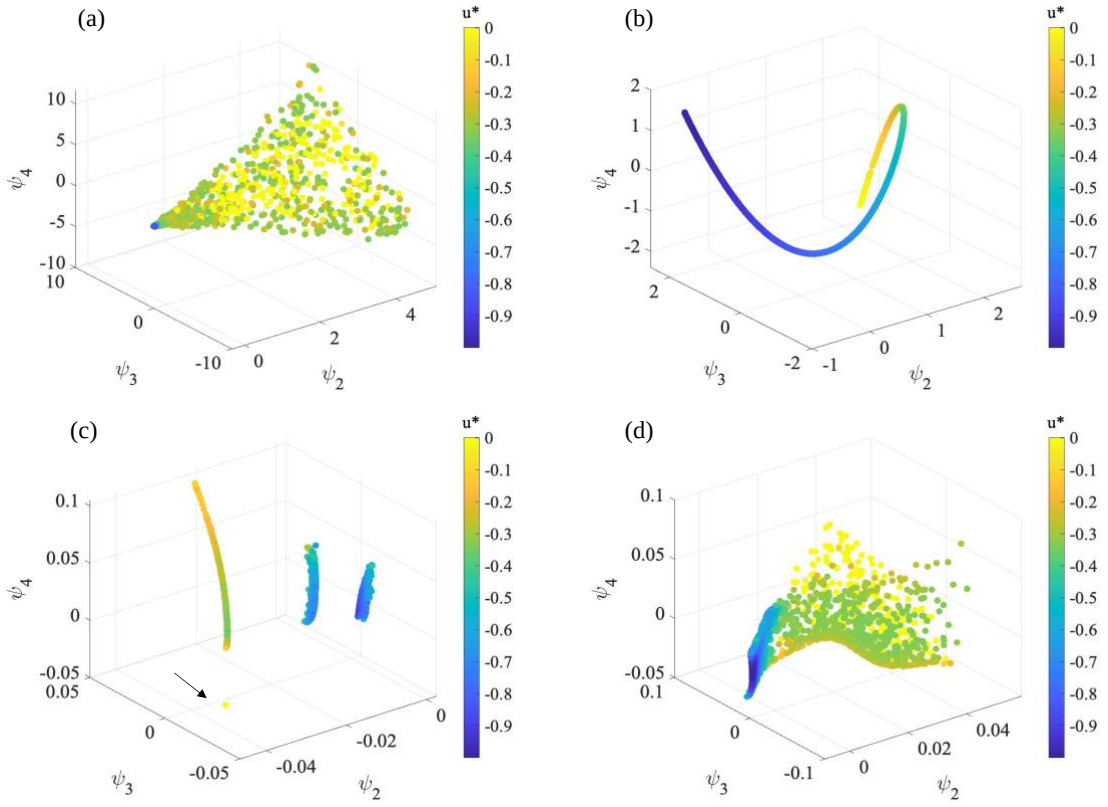

Figure 2.11. Subspaces generated by Diffusion Maps for LJ3 at T^* = 0.18. (a) Hausdorff Distance (b) Mayer *f*-bond (c) IsoRank (d) Spectral Distance. Data is colored based on potential energy.

In Figure 2.11(a), DMap results, which are obtained by using the Hausdorff distance metric, manifest that the first non-trivial eigenvector (ψ_2) can effectively discriminate the configurations having lower potential energy values (blue region) from the ones with the moderate (green region) and high (yellow region) energy ones. On the other hand, this eigenvector cannot discriminate between configurations with moderate and high potential energy values. The third eigenvector has no contribution to discriminate the data further than what the second eigenvector has already done, and essentially this eigenvector seems to be capturing only a symmetry effect. These outcomes were expected, considering the nature of the Hausdorff distance, which is based on the maximum distance of the particles in one configuration from the closest particle in the other

configuration. Thus, it is not good at discriminating between any two 'broken' structures, which are represented in Figure 2.3(b). The segregation between the moderate and high levels of potential energy, which could not have been previously accomplished by the Hausdorff Distance, can be successfully presented by the Eigenspectra distance metric (Fig. 2.11(d)). Potential energy is well-correlated by the corresponding eigenvector in the continuous subspace. Different configurations in completely dissociated clusters can be captured with this distance metric. It can also be viewed as a continuous form of the IsoRank-based subspace with the continuity arising due to the use of a continuous variable to represent the bond distance between two particles. Another key difference is that instead of collapsing all fully broken clusters to one point, the Eigenspectra distance captures the different modes within the entirely broken configurations. On the other hand, Figure 2.11(c) shows that employing the IsoRank distance metric indeed yields in a result in which the data are separated into four distinct clusters, previously verified as corresponding to the four structural motifs shown in Figure 2.3(b) (this will be exhibited more clearly under the order parameter analysis in section 2.4.4). The correlation of the individual diffusion coordinates with the potential energy values is weak, but the energy distribution within each data grouping is quite smooth. The IsoRank-based diffusion mapping reduced the fully disconnected configurations to a single point, as the unbonded particles are clipped away by the multiplication with zero. Detection of the mentioned point in the plot is somewhat challenging; nevertheless, its exact location was guided by an arrow. Figure 2.11(b) shows subspace that were obtained by using Mayer *f*-bond definition to create distance metric. It was concluded from the results that only one dimension is enough to describe the system features since there is a strong correlation between the non-trivial second eigenvector and the potential energy of the system.

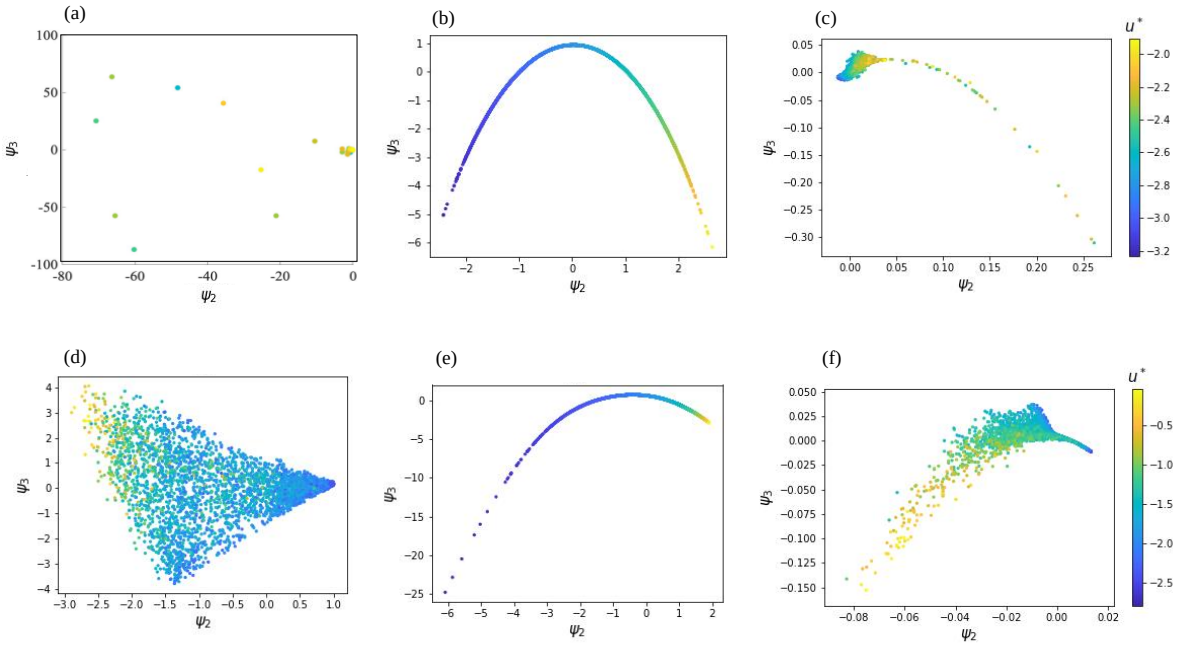

Figure 2.12. Subspaces generated by Diffusion Maps for LJ13 at T^* = 0.28 (top) and T^* = 0.4 (bottom). (a) and (d) Hausdorff Distance (b) and (e) Mayer *f*-bond distance (c) and (f) Eigenspectra Distance. Data is colored based on potential energy.

For the LJ13 system, subspaces created with the eigenvectors for two different sampling temperatures can be found in Figure 2.12. For these systems, the same distance metric types are investigated except the IsoRank since it is intractable for a higher number of particles, as was also indicated by Long and Ferguson (2014). The data for T^* = 0.28 and T^* = 0.4 in the reduced coordinates is represented at the top and the bottom of Fig. 2.12, respectively. The primary investigation proved that the data could be embedded in two dimensions; therefore, diffusion coordinates are represented in 2-D instead of 3-D. Data are colored based on the corresponding potential energy of the configurations in order to see the possible correlations. DMap analysis, which has been performed using Hausdorff distance, showed (Fig. 2.12(d)) very similar results with LJ3 systems for the T^* = 0.4 case, where the second eigenvector can capture the information about the structure of configuration whether it is broken or clustered. However, this distance metric cannot succeed to segregate different types of the broken structures, as there is no clear distinction

between the blue and green regions in Figure 2.12(d) with respect to the second eigenvector. Results for T^* = 0.28 (Fig. 2.12(a)) cannot discriminate the structural differences in the configurations. Therefore, instead of the continuous distribution throughout the diffusion coordinates as a function of u^*, only discrete points can present in which they have clustered for similar potential energy values. Figures 2.12(b) and 2.12(e) draw the same conclusion about the results of the DMap obtained via Mayer *f*-bond distance metric for two different temperatures with the results of the three-particle system. Independent from the number of particles, there exists a perfect correlation between u^* and second eigenvector, which is suggested by DMap analysis as one and only variable that can describe the system characteristics. Although this distance metric perfectly segregates all different configurations based on the corresponding potential energy values, the need of having the potential energy data is an important limitation of this type of distance metric, since the underlying potential energy function may not generally be known for a given empirical data set. This was expected based on the definition of this distance metric, as discussed in section 2.3.1.1 above. In practice, this metric is not directly useful for the problem at hand, but it does demonstrate that the DMap technique generates a result that is consistent with the distance metric employed. Figure 2.12(c) shows the subspace generated and colored by u^* for the LJ13 system simulated at T^* = 0.28 for the Eigenspectra distance metric. u^* does not show a strong correlation with the diffusion distances, which is expected since within the manifold itself, u^* did not previously exhibit any correlation within the Cartesian coordinates (Fig. 2.7 (b). Nevertheless, for reasonably higher u^* values, the plot shows a slight correlation with respect to the second eigenvector. A similar outcome can be observed for T^* = 0.4 cases as well (Fig. 2.12(f)). Different than T^* = 0.28, the second eigenvector relates better with u^*, and a broader trend was observed due to the nature of u^* itself, which has a broader spread in u^* values, in

comparison to the lower temperature case. This broader range of u^* values could be attributed to a greater variety of configurations that are seen at the higher temperatures, as represented in Figure 2.4(a).

2.4.4 Investigation of the Order Parameters

In section 2.4.3, it was demonstrated that the DMap could successfully identify the low-dimensional picture of the data. Nevertheless, this method cannot provide direct information about the physical significance of a new variable space. For the classification of the set of order parameters that are computable for a given configuration solely from the information in the original coordinate space $\boldsymbol{X}$, a heuristic study has been conducted.

In order to obtain the variables that can describe the system characteristics and can serve as substitutional variables to explain the thermodynamics of the system, an extensive study has been performed with several order parameters. The radius of gyration is one of these parameters, and it was previously extensively studied (Beltran-Villegas & Bevan, 2011; Beltran-Villegas et al., 2011; Beltran-Villegas et al., 2012). Other intuitive variables, including the number of the bonds n_b or connectivity C, are relatively unstudied and investigated as the order parameters in this study.

It is important to note that investigations of the order parameters based on the relation with the non-trivial eigenvectors have been employed only for IsoRank and Eigenspectra distance metric results. Mayer f-bond has already exhibited an excellent association with u^*; therefore, there is no need for this distance metric to be correlated with any order parameters. For Hausdorff distance metric, on the other hand, it will not be possible to capture any discrepancies between different forms of the broken structures based on its nature. Hence, in the rest of this section, eigenspectra distance and IsoRank will be used as the primary screening tools to find out the order parameters, based on what we have learned about the nature of the method.

Figure 2.13 shows the order parameters that have been found after the heuristic study for the LJ3 system. The figures are colored by two structural order parameters as n_b at the top figures and r_g at the bottom figures. Figure 2.13(a) shows that n_b can identify the structural differences between different broken cases where n_b has low values, in the direction of ψ_2. The results for the IsoRank-based distance metric (Fig. 2.13(b)) look qualitatively different because this metric creates a discrete grouping of data in the eigenvector space. Furthermore, n_b can correlate well in a discrete form with the first two non-trivial eigenvectors. The variable r_g describes changes in shape within a given bonding state. Figure 2.13(c) shows that r_g correlates very well with ψ_2 across the entire range for the eigenspectra distance. It is not the same with IsoRank distance (Fig. 2.13(d)) metric based on its discrete nature; this distance metric might not be used to be described by a continuous structural variable like r_g.

Figure 2.14 depicts the correlation results of the first two non-trivial eigenvectors ψ_2 and ψ_3 for the LJ13 system. As stated earlier, for the sake of a better physical explanation for the higher number of particles, instead of the number of bonds (n_b), connectivity (C) is used as a measure to capture the structural changes in the configurations. Figure 2.14(a) shows the subspace generated with the eigenspectra distance metric and colored by r_g for LJ13 simulated at $T^* = 0.28$. It is clearly observed that the second diffusion coordinate ψ_2 encodes r_g very well, while ψ_3 cannot capture any symmetry variable. However, as mentioned before in the context of Figure 2.11(c), within the manifold itself, ψ_3 encodes the spread of u^*. This implies that the diffusion coordinates generated by the eigenspectra distance metric encode r_g entirely in ψ_2 with ψ_3 encoding the energy distribution at a specific value of r_g. This can be said for the subspace generated for $T^* = 0.4$ as the subspace (Fig. 2.14(b)) showed a similar spread of u^*, as was previously seen in Figure 2.11(b).

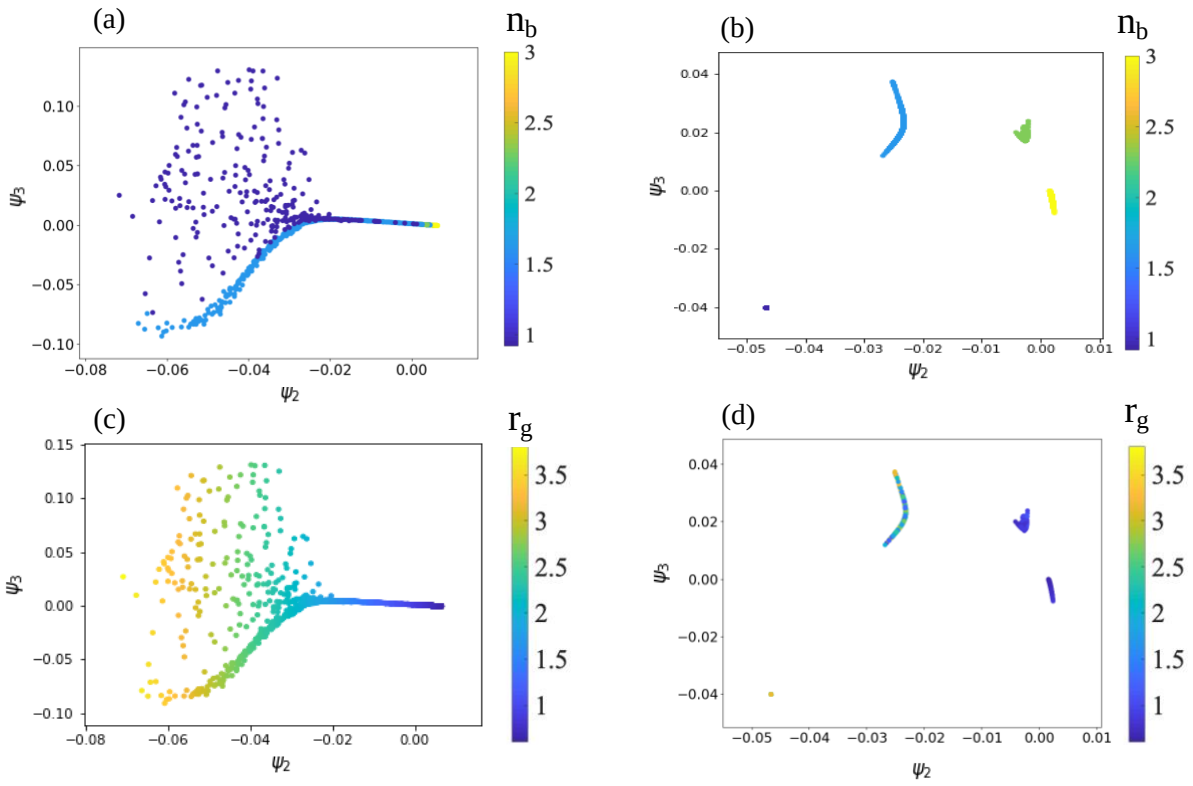

Figure 2.13. Subspaces generated by DMap colored by the variable yields to the correlation for LJ3. (a) and (c) Eigenspectra distance metric. (b) and (d) IsoRank distance metric. Top figures are colored by n_b and bottom figures are colored by r_g.

For very low values of r_g, the contribution of the third eigenvector in $T^* = 0.28$ is significant, whereas the value of r_g increases, only ψ_2 can describe the system. For $T^* = 0.28$, on the other hand, the contribution of the ψ_3 is observed for the high values of r_g. This difference can be attributed to the nature of the dataset. The dataset at $T^* = 0.28$ has the majority of the data with low r_g values; therefore, a few data points are presented with high r_g values. Whereas, the dataset at $T^* = 0.4$ has a variety of r_g values as the energy is very high, and therefore particles in the cluster show less bonding formation. In order to find the contribution of ψ_3, correlations with C are investigated for both T^* values. Figure 2.14(c) shows that ψ_3 correlates well with C for its low to intermediate values. However, the same conclusion cannot be drawn for $T^* = 0.4$, as can be seen

in Figure 2.14(d) since there is no correlation observed between ψ_3 and C for the region where the data shows more spread distribution.

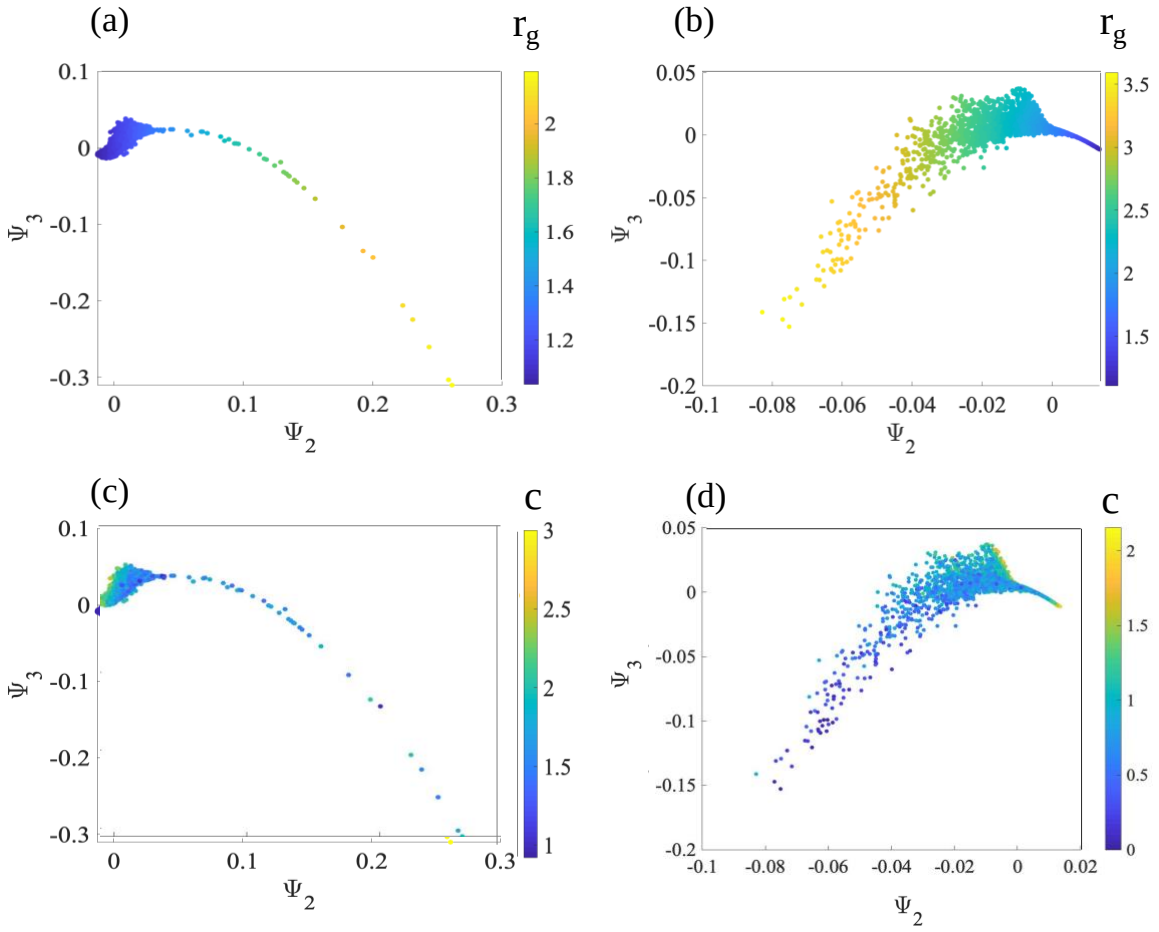

Figure 2.14. Subspaces generated by DMap colored by the variables yield to the correlation for 13-particle system. T^* = 0.28 (left), T^* = 0.4 (right). Eigenspectra Distance metric colored by (a) and (b) r_g, (c) and (d) C

Based on these results, r_g, and n_b (or C), will likely provide a good set of parameters for both LJ systems. Usage of these variables as the new thermodynamic variables can be studied by the aid of Figure 2.15, which demonstrates the complete picture of the functionality between the variables and the potential energy. Figure 2.15(a) shows that n_b divides the data into four main structural motifs, meanwhile r_g is responsible for capturing the different modes of the motions. Small changes in r_g cause substantial changes in u^*, and this yields to the stiffest case, where the

value of n_b is three. Although the data shows a broad behavior for the case $n_b = 1$, which could be credited to the modes of motion that r_g cannot capture; overall, the collapse of the data was successfully executed. Similarly, Figures 2.15(b) and 2.15(c) depict the potential energy landscape for a 13-particle system in the reduced space. The variables discovered are the same as the three-particle system, but as clarified before, C has been used as the counterpart of the number of bonds not to deal with the scaling for large values of n_b. For $T^* = 0.28$, there is a small spread in the dataset for the low values of potential energy, which is reasonable since both r_g and C show some spread at that range of potential energy as can be clarified with their individual correlation plots in Figure 2.15.

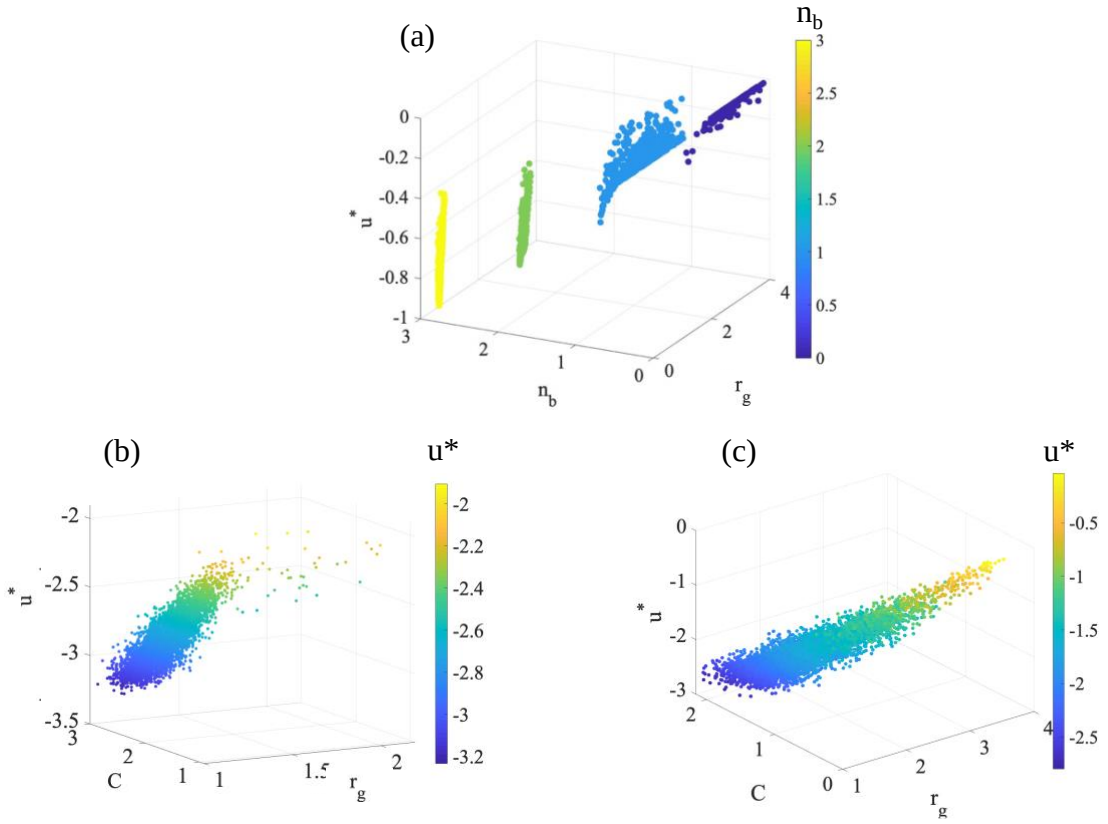

Figure 2.15. Energy Landscape with reduced parameters. (a) $N = 3$, $T^* = 0.18$ (b) $N = 13$, $T^* = 0.28$, (c) $N = 13$, $T^* = 0.4$.

After the relation between u^*, r_g, and C has been obtained to show the analytical equation of u^* as a function of C and r_g, a regression analysis has been performed by using MATLAB non-linear multi-dimensional regression function. The model equation is given in Equation 2.20, and the corresponding coefficients can be found in Table 2.1.

$$u^* = \left(a_1 * \left(r_g - a_2\right)^2\right) + \left(a_3 * \left(r_g * C\right)\right) + \left(a_4 * (C - a_5)^2\right) + a_6 \tag{2.20}$$

Table 2.1. Fitting coefficients for Equation 2.20

T*	a_1	a_2	a_3	a_4	a_5	a_6
0.28	-2.7611	1.5281	0.3481	0.0705	6.3099	-4.4123
0.4	0.0059	-35.7733	-0.089	0.1328	3.1038	-10.5523

The regression coefficient R^2 was calculated as 0.796 and 0.887 for $T^* = 0.28$ and $T^* = 0.4$ datasets, respectively. The points with high potential energy values, which are distinguished by the spread in the dataset, reduces the value of the R^2 for $T^* = 0.28$.

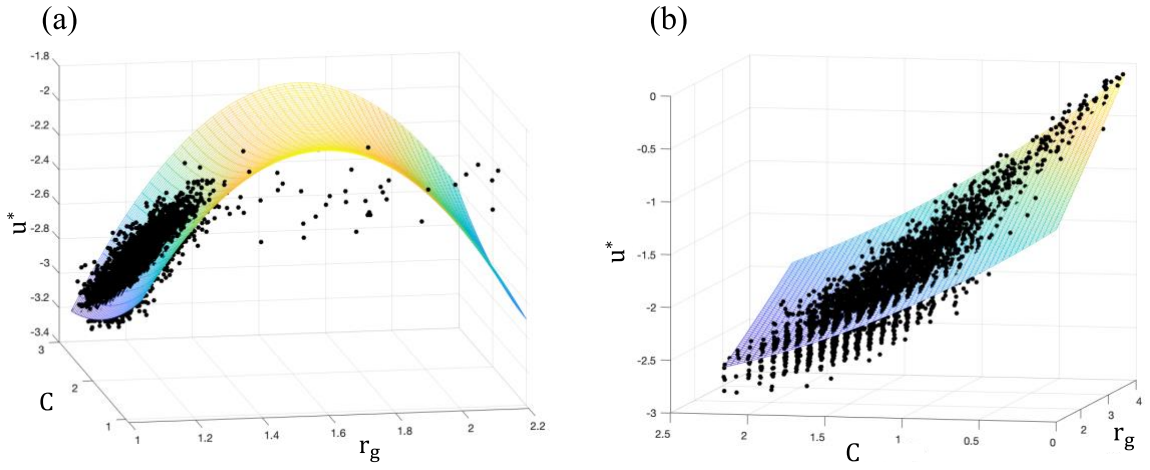

Figure 2.16. Combination of the results of the model equation and the original datasets for the LJ13 system. (a) $T^* = 0.28$, (b) $T^* = 0.4$. The fitted surface is the model equation results, and black points are the original data points.

The fitted surfaces which represent the model equation and the dataset at the corresponding temperatures can be seen in Figure 2.16. Considering the good agreement between the model equation and the datasets, using this equation as a tool to interpret the system's thermodynamic state can be suggested.

2.5 Conclusions

In this study, dimensionality reduction was applied to two different systems, having a different number of particles (three and 13) that are interacting via Lennard-Jones potential. Datasets for both systems were created by canonical ensemble Monte Carlo simulations. DMap analyses concluded that, for both three and 13 particle systems, the datasets were found to be successfully interpreted with two independent variables, rather than the initial number of dimensions of nine and 39, respectively.

The data collected for a 3-particle system at $T^* = 0.18$ exhibits the most structural variation. For the LJ13 system, T^* values are selected as 0.28 and 0.4, considering the transitions between solid-like liquid-like structures and the liquid-like gas-like structures based on the observations. In order to fully exploit the characteristics of the data, and describe the thermodynamic relations of a small system, the quantity, as well as the types of independent variables related to the system have to be evaluated. As mentioned earlier, DMap analyses revealed that the number of variables that can describe the system was found to be two, rather than nine, which signals a complex fashion. It has also been shown that by taking heuristics into consideration as well; the thermodynamic state can be completely explained when the potential energy is a function of the two reduced variables, one being the discrete and other continuous which are the number of bonds n_b and the radius of gyration r_g for the current system. Moreover, the results can be used to create an analogy with a moderately small system, where the potential energy is

traditionally described as a function of volume and area. Hence, it can be concluded that the outcomes gained from this simple system of $N = 3$ can open pathways to analyzing larger systems with high complexity without a dramatic increase in the number of variables required. On the other hand, for the LJ13 system, DMap was found to be effective in reducing the dimension of this system as well, despite the noisy nature of the complex dataset. The number of variables was decreased from 39 to two for this system, where those two effective variables were found to be connectivity C and radius of gyration r_g. After the order parameters are detected, potential energy is modeled as a non-linear function of C and r_g using the regression.

Overall, high dimensional Monte Carlo simulation data was used to explore the applicability of machine learning techniques on dimensionality reduction. Precise recognition of distorted motifs in high-dimensional Lennard Jones cluster systems was achieved by using diffusion maps, and a sufficient number of reduced variables that can successfully describe the complex systems was detected. Future works could explore detecting parameters in more complex systems that could act as state variables on which an energy landscape can be built on, to achieve a more controlled design of experiments in research which are based on colloidal and interfacial sciences.

CHAPTER 3

PHASE TRANSITIONS IN BULK SQUARE WELL AND HARD DISK BINARY MIXTURES

3.1 Introduction

Phase transitions have long been a topic of interest since they play a role in design of materials. Hence, predicting the phase transitions is crucial in analyzing the material for further development with the macroscopic properties. Phase transitions can be studied experimentally, theoretically, or by the aid of computational studies. This study focuses on the computational side of the phase behavior examinations. Working with simulations leads to a convenience for the cases where executing an experimental study of the system is challenging, or sometimes even impossible. Moreover, to be able to predict the possible phase diagram based on molecular interactions for any system could be helpful when the theory needs to be validated, or for the cases where the applicability of the theory onto the studied system is not possible.

Throughout the years, to be able to create basic model systems in the computational environment, one of the most commonly used types of pair potential to model the particles is the hardcore potential (u_{hd}). This fundamental and simple approach to evaluate the potential between particles solely describes repulsion in short distances without any attractive interactions. Hardcore potential, namely hard-sphere in three-dimensions and hard disk in two-dimensions, which is the focal area for this study, is shown in Figure 3.1 and described in Equation 3.1.

$$u_{hd}(r) = \begin{cases} \infty & r < \sigma \\ 0 & r \geq \sigma \end{cases} \tag{3.1}$$

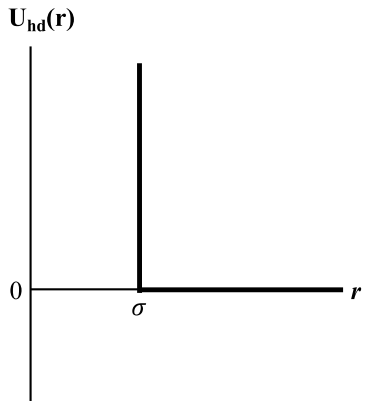

Figure 3.1 Representation of the hard disk potential where σ is the characteristic diameter of the particles, and *r* is the distance between the centers of the particles.

Another simplistic model is the square well potential, which also takes the attractive interactions between the particles into account in addition to the repulsion. Basically, for particles having the distance smaller than σ there exist repulsive interactions. However, the interaction between those particles is attractive with strength ε and characteristic range (δ) outside the repulsive core (Wood 1970; Wood and Erpenbeck 1976). The square well pair potential is defined in Equation 3.2, and the corresponding representation can be found in Figure 3.2.

$$u_{sw}(r) = \begin{cases} \infty & 0 \le r < \sigma \\ -\varepsilon & \sigma \le r < \sigma + \delta \\ 0 & r \ge \sigma + \delta \end{cases} \tag{3.2}$$

where ε is the square well depth.

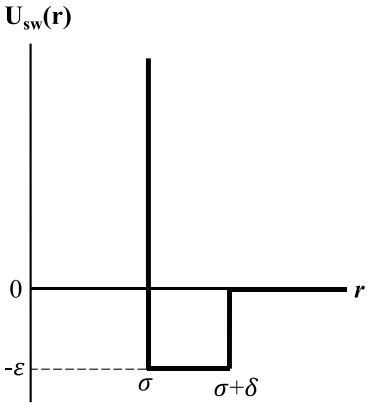

Figure 3.2. Representation of the square well potential

Through the decades, both pure component square well and hard disk potentials were extensively studied, because of their highly applicable properties to the real-life experiments, and intriguing phase transition characteristics that they exhibit. Additionally, the characteristics of square well mixtures and hard disk mixtures have also been well-documented due to the mentioned qualities. Although individual examinations of hard disk and square well potentials are considered primitive and well-examined, in terms of the principles behind the phase behavior of simple and atomic structures, to our knowledge, the mixture of the square well and hard disk particles has not been extensively studied. Therefore, the phase transition behavior of this binary mixture is still not well discovered and needs further clarification. Here, it is noteworthy to clarify that binary mixture denotes a system whose components are identical in terms of the size and shape but differ in the type of potentials they possess.

The biggest motivation behind this study is the observation of the isostructural solid-solid transition due to the short-range interaction, which is first reported by Bolhuis and colleagues (Bolhuis, Hagen & Frenkel 1994). Isostructural phase transition is a symmetry preserving, and it resembles a first order liquid-solid transition, which ends at the critical point. It was shown that both two and three-dimensional systems that are simulated with square well potential undergo a solid-solid phase transition when the ratio between the attraction potential and particle diameter is less than 0.07. Moreover, this isostructural solid-solid phase transition, which is caused by the short-ranged attractions, can also be encountered in real applications such as uncharged colloids, nonadsorbing polymers, and hard disk colloid mixtures. Given in the mentioned paper, the phase diagram of the pure square well particles which undergo an isostructural solid-solid phase transition for the δ range of 0.01 to 0.07 can be found in Figure 3.3.

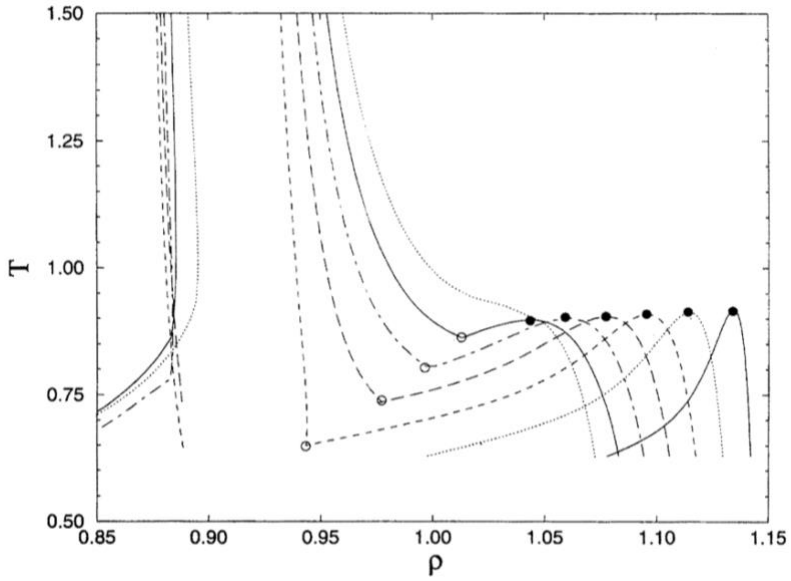

Figure 3.3. Phase diagram of a two-dimensional pure square well particle system. The range of attraction (δ) increases from 0.01 (far right) to 0.07 (far left) for the solid-solid transition curves. (Source: Bolhuis, Hagen & Frenkel, 1994).

Phase diagrams could be constructed via theoretical methods, such as fundamental measure density functional theory (Smithline & Haymet, 1987; Likos, Nemeth & Lowen, 1994), thermodynamic integration(van Gunsteren and Berendsen 1987; Elliott and Hu 1999). Gibbs-Duhem integration (Kofke 1993) and perturbation theory (Rosenfl & Thieberger, 1973; Chang & Sandler, 1994; Velasco, Navascue & Mederos, 1999); or with the simulations of Gibbs ensemble (Vega et al., 1992; Giacometti et al., 2009). Parameters of the phase diagram can be evaluated either with the perturbation theory followed by Gibbs-Duhem integration or implementing Maxwell construction on free energy. In this study, coexistence densities are computed with the double tangent construction method, which shares the same underlying idea conceptually with the Maxwell construction, yet more preferable for the systems working with Helmholtz free energy instead of Gibbs free energy. Although this study has focused on the binary mixtures, the equation which was proposed for the pure square well solids with the new set of coefficients was adapted into the free energy calculations (Bolhuis, Hagen & Frenkel, 1994). Further details about the

computations can be found in section 3.3. Additional comments about the alternative calculation route were given at the end of the results and discussion section.

The main objective of this study is to create a phase diagram of the hard disk and square well binary mixture by the aid of free energy calculations. For that purpose, the pure square well solid-solid isostructural phase diagram is prepared, and the effect of the additional hard disk component on the phase diagram is investigated simulating for 75% square well %25 hard disk binary mixture for a fixed attractive interaction parameter δ as 0.04.

3.2 Literature Review

Understanding the thermodynamic characteristics of the substances through their phase transition behaviors has gained a substantial attraction to lead researchers to work in this field. Phase transition behavior of various material systems have been extensively studied thus far; however, in this section, particular attention has been given to the studies that are conducted mostly by molecular simulations. Existing studies show that molecular simulations with different model potentials, including hardcore, square well, or Yukawa potential, can be directly consolidated with the upcoming computational or experimental studies that aim to create the broad new approaches in condensed matter physics and colloidal science. Therefore, these studies can navigate both theoretical and practical advancements in the near future. As the main focus of this study is obtaining phase diagram of square well and hard disk binary mixtures, a collection of the existing studies that concentrate on the mentioned types of potentials are briefly summarized in the following paragraph.

The pure hardcore potential has been widely studied, and their characteristics are well documented. Nevertheless, the nature of the potential varies in two-dimensional hard disks and three-dimensional hard spheres, thus creating an exciting demand to examine the phase transition.

An early work about the hard disk potential has been performed via molecular dynamics simulations for 870 hard disk particles by Alder, and Wainwright (Alder and Wainwright 1962) concluded that an increase in the density causes a phase transition from liquid to two-dimensional crystal solid structures. Based on the ability of hard disk particles to provide a geometric description of freezing at the packing fraction of 0.69, they have been primarily modeled in the studies that mainly focus on the statistical mechanics of the liquid systems (Monson & Kofke, 2007). After this remarkable results presented by Alder and Wainwright (1962), freezing (and melting) processes in two-dimensional hard disk particles were continued to be investigated to fully comprehend if this phase transition is a well-known first-order transition, or some other continuous transition form (Alder, Hoover & Young, 1968a; Mitus, Weber & Marx, 1997; Wood 1968; Erpenbeck & Luban, 1985; Da Silva, Caliri & Mokross, 1987; Ree & Hoover, 1964; Truskett et al., 1998; Huerta, Henderson & Trokhymchuk, n.d.; Kawamura, 2005; Hu & Yu, 2009; Bernard & Krauth, 2011). Moreover, Speedy (Speedy, 1999) provided the thermodynamic features of the glass transition in binary hard disk mixtures, which were composed by a ratio of the particle diameters is approximately 0.77. Huerta and colleagues presented packing fractions for binary hard disk mixtures with a range of particle diameters varying from one to 1.4 in order to comprehend the difference in the freezing between pure hard disk particles and binary mixtures. Their results claimed that the addition of another hard disk component with a different diameter moves the freezing transition point to high packing ratios (Huerta, Carrasco-Fadanelli & Trokhymchuk, 2012).

The square well potential was utilized in a vast amount of studies either computationally (Brilliantov & Valleau, 1998; Orkoulas & Panagiotopoulos, 1999; Armas-Pérez et al.,2014), or theoretically(Henderson, 1974; Bokis & Donohue, 1995) as a fundamental approach to model the

intermolecular attractions of the particles. Additionally, it gives an adequate explanation for the uncharged colloidal particles with a short-ranged attraction. As mentioned previously, square well particles exhibit features of attractive forces that real physical systems possess. The addition of short-range interactions, whether it is attractive or repulsive, greatly impacts the high-density phase behavior. It can also cause symmetry conserving transition between two crystal phases. Isostructural phase transition was first reported by Young and Alder (Young & Alder, 1979) by using a square well potential with long-range attractions. They concluded that the results were not very realistic based on the long-range attraction that they have used to model the particles. The phase transition from condensed to expanded solid phases between the particles with a face-centered cubic (fcc) structure has been studied by Bolhuis and Frenkel (1994). The intermolecular attraction was approximated based on the square well potential that can be related to the uncharged colloidal particles with short-range attraction. For the particles that have sufficiently short-ranged interactions, a phase transition has been found to be occurring parallel to phase change between liquid and vapor phases with a long-range attraction, since it ends in critical temperature. Another computational study to describe the phase behavior of a hard-sphere system with attractive Yukawa interactions has been done by Dijkstra (Dijkstra, 2002). In the same study, spinodal instabilities in solid and fluid phases, along with the metastable isostructural phase transitions between the fluid and solid phases, have been discussed. Phase diagram of square well systems with a short-range attraction has also been investigated by Likos and Senatore (Likos & Senatore, 1995) and the isostructural phase transition between stable crystalline phases of particles was observed to be in a similar trend with the aforementioned studies. Results obtained by density functional theory (DFT) showed that the observed phase transition is isostructural when the ratio of attraction potential to the particle diameter is less than 0.05. Also, increasing the said ratio was

found to be triggering a melting. The particles with short-range attractive square well potential are explored by Foffi et al. (2002). In this study, phase transitions have been investigated with the variation of potential range, and interaction diameter to repulsive core to provide theoretical information to the several application areas such as nanoparticle aggregation and protein crystallization. Their results showed that as the range of potential parameter decreases, critical point shifts from lower density to higher density.

Apart from the studies that investigate pure square well particles, the focus on the component mixtures of square well systems (Wang & Wang, 1990; Gulati & Hall, 1997; Maldonado et al., 2015) has enlightened the field, particularly the colloids science. The outcomes of such studies serve significant purposes by creating unique ideas for real-life applications as the thermodynamic behaviors of these systems are exploited (Alder, Alley, and Rigby 1974). Similarly, existing literature has a variety of studies about one component mixture of hard disk particles (Speedy 1999; A. Huerta, Carrasco-Fadanelli & Trokhymchuk, 2012; Russo & Wilding, 2017; Lin & Oettel, 2018) and their three-dimensional counterparts (Kranendonk & Frenkel,1989; Denton & Ashcroft, 1990; Eldridge, Madden & Frenkel, 1993; Velasco, Navascué & Mederos, 1999; Warshavsky & Song, 2008; Kummerfeld, Hudson & Harrowell, 2008).

As one of the common real-life examples, colloidal systems have been the focal point of various studies to show the isostructural solid-solid phase transitions. Colloids can interact with an effective potential with an attractive minimum with the addition of non-adsorbing polymers. More recently, a theoretical study has been conducted by Garcia and co-workers to predict the phase behavior of colloid superballs and non-adsorbing polymer dispersion with experimentally observed phase behaviors (García et al., 2018). In their work, superballs and non-adsorbing polymers are modeled as hard spheres. The results that were obtained by using free volume theory-

a commonly used method to explain the phase behavior of colloid dispersions- showed that the said mixture exhibits an isostructural phase transition in both fluid-fluid and solid-solid cases. Coexistence between simple cubic phases was found to be dependent on the curvature of the particles, and the particles having more curvature need a higher depletant concentration in order to be in the coexistence between phase transitions. For the long-ranged depletion attraction cases, there exists an isostructural fluid-fluid coexistence irrespective of the shape parameters. As a perfect reference system for colloid-polymer mixtures, a mixture of hard spheres, which were used as depletants, was studied, and isostructural solid-solid phase transition was perceived experimentally when the size ratio between depletant and colloid is smaller than 0.091 (Garcia et al., 2018). Rey and co-workers experimentally observed the isostructural solid-solid phase transition between two crystalline phases at core-shell microgel-laden fluid interfacial monolayer. By utilizing the serial monolayer transfer procedure, notable challenges at the experimental procedures have been overcome, and a proper experimental system leads to study two-dimensional system which has an interaction at contact due to the repulsion with the length scales (Rey et al., 2016).

3.3 Computational Methods

This section contains the details of data generation and the calculation with the related equations used in the course of calculations. First, the simulation model and the details are described. Next, thermodynamic relations in order to compute the free energy of pure square well and hard disk square well binary mixtures are given. At the end of this section, the double tangent construction method, which is used for obtaining phase coexistence densities, is explained.

3.3.1 Simulation Details and Model System

In this study, the canonical ensemble Monte Carlo simulations have been used to find the thermodynamic variables of the system at varying volume and temperature values. The studied case includes 196 particles in the square box. For the computations in Monte Carlo, periodic boundary conditions are applied to the system. The acceptance ratios for the translational and rotational movement of the particles are set to 20% by varying the δr_{max} whose value was initially given as 0.1σ. After examining the fluctuations in the energy for varying combinations of the number of production steps, the total number of steps for each run is employed as 10_7, with the number of the equilibrations and number of passes per subaverage as $5x10_4$ and 10_4, respectively. For each run, the initial configuration is created as a hexagonal lattice for the given temperature and the volume. In the binary mixture case, the ordered lattice structures are designed in a way that a unit cell would have one hard disk particle at the corner, and square well particles are located at the other three edges as can be visualized in Figure 3.4.

The density and temperature range for both pure square well cases and square well and hard disk binary mixture is determined similarly to the reference study. The range of reduced density is defined between 0.8 and 1.14 with a spacing of 0.1, and the reduced inverse temperature (ε) is set in the range of 0.1 to two with the grid size of 0.1. A total of 660 sets of combinations for temperature and density are needed to be run to collect energy data for the pure square well and the binary mixture cases separately. After the energy data is obtained as the output of each simulation run, corresponding free energy values are calculated to evaluate the coexistence density and temperature values to generate the phase diagram. In the following part, the details about the free energy computations are given.

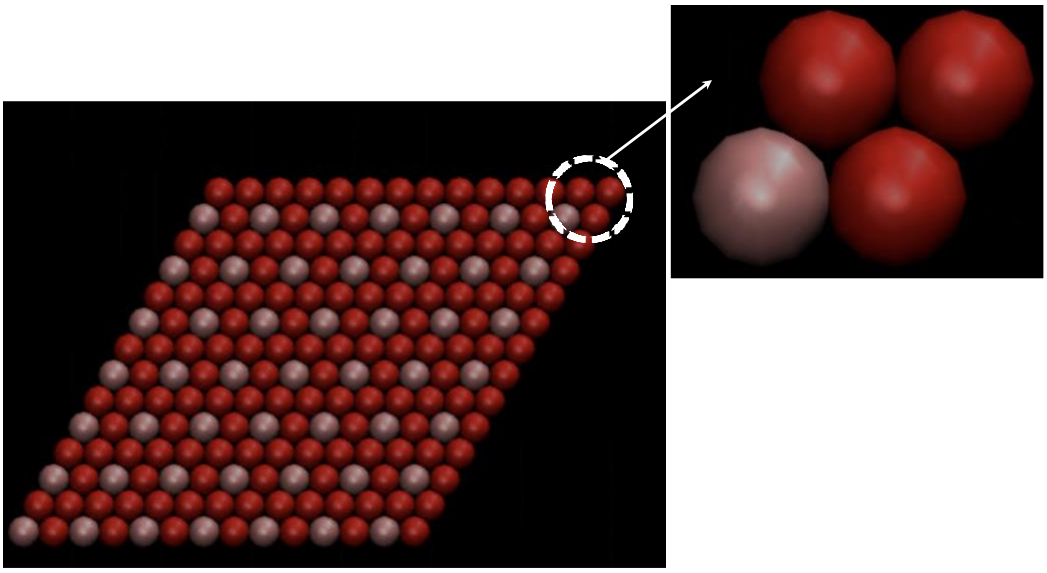

Figure 3.4. Representation of the (a) initial configuration snapshot for T = 0.86, ρ = 1.102, (b) enlarged version of the unit cell with one hard disk (pink) and three square well (red) particles

3.3.2 Computation of the Free Energy

The free energy of a square well solid phase cannot be evaluated directly by using canonical ensemble Monte Carlo simulations. To calculate the free energy values as a function of density and temperature, thermodynamic integration (Frenkel, 1991) is performed. To be able to create a reversible path, the free energy of a hard disk calculated for the corresponding density values are used as the reference state with the equation given below.

$$F(\rho,\varepsilon) = F_{HD}(\rho) + \int_0^{\varepsilon^*} \frac{\langle U\rangle_{\varepsilon^*}}{\varepsilon^*} d\,\varepsilon^* \tag{3.3}$$

where $\varepsilon^*=\frac{\varepsilon}{k_B T}$ as k_B is being Boltzmann constant, and T is the temperature. $\langle U\rangle_{\varepsilon^*}$ is the canonical ensemble average energy obtained at constant ε^*. $F_{HD}(\rho)$, and $F(\rho,\varepsilon)$ are the free energy of hard disk and square well solid, respectively. Bolhuis and co-workers (Bolhuis, Hagen & Frenkel, 1994) have suggested using the relation between $\langle U\rangle_{\varepsilon^*}$ and the number of atomic pairs, N_p, and the mentioned relation is stated below.

$$\langle U \rangle_{\varepsilon^*} = N_p \varepsilon \tag{3.4}$$

Since this equation was originally used for the pure square well solid particles, the values of $\langle U \rangle_{\varepsilon^*}$ were obtained from the pure-square well computer simulations. However, as the context of the current study, a pseudo-pure approach was taken where binary mixtures were treated as pure components, and the results were evaluated.

By modifying Eq. 3.3 by substituting the $\langle U \rangle_{\varepsilon^*}$ term in Eq. 3.4, and dividing each side by Nk_BT, the final equation which gives the free energy is formed.

$$\frac{F(\rho, \varepsilon)}{Nk_BT} = \frac{F_{HD}(\rho)}{Nk_BT} - \int_0^{\varepsilon^*} \frac{\langle N_p \rangle}{N} d\varepsilon^* \tag{3.5}$$

The term F_{HD}, which is the abbreviation of the hard disk free energy, is calculated with the equation (Eq. 3.6) given in the study by McBride & Vega, (2002). It is noteworthy to mention that, although the mentioned study focuses on the evaluation of the hard disk dimer case, the equation related to the monomer case was also included and is used in our research.

$$\frac{F^{total}}{Nk_BT} = \frac{F^{ideal}}{Nk_BT} + \frac{F^{residual}}{Nk_BT} \tag{3.6}$$

In Eq. (3.6), F^{ideal} and $F^{residual}$ are the ideal and residual components of Helmholtz free energy of a hard disk, F^{total}. The individual definitions of F^{ideal} and $F^{residual}$ are given in Eq. 3.7 and 3.8, as follows.

$$\frac{F^{ideal}}{Nk_BT} = ln(\rho\sigma^2) - 1 \tag{3.7}$$

$$\frac{F^{residual}}{Nk_BT} = \frac{F^{residual}(\eta_0)}{Nk_BT} + \int_{\eta_0}^{\eta} \frac{Z^{solid} - 1}{\eta} d\eta \tag{3.8}$$

where, η is the volume fraction and for the 2-dimensional particles, its formula is given in Eq. 3.9.

$$\eta = \rho \frac{\pi}{4} \tag{3.9}$$

$F^{residual}(\eta_0)$ is given as 3.818 for $\eta_0 = 0.716$ (McBride & Vega, 2002). Equation of state for the solid phase, Z^{solid} is given in Equation 3.10 (Alder, Hoover, & Young 1968).

$$Z^{solid} = \frac{2}{\alpha} + 1.90 + 0.67\alpha \tag{3.10}$$

where α is expanded as,

$$\alpha = \frac{\rho_0 - \rho}{\rho} \tag{3.11}$$

Close packing value ρ_0 can be found in Eq. 3.12.

$$\rho_0 = \frac{2}{\sigma^2\sqrt{3}} \tag{3.12}$$

Consequently, the result of integration becomes,

$$\int_{\eta_0}^{\eta} \frac{Z^{solid} - 1}{\eta} d\eta = -\frac{0.6077}{\eta} - 2\log(0.9069 - \eta) + 0.23\log(\eta) \tag{3.13}$$

With the substitution of $F^{residual}(\eta_0 = 0.716)$ in the integration given in the Eq. (3.13), hard disk free energy can be written as in the following form.

$$\frac{F_{HD}(\rho)}{Nk_BT} = 2.818 - \frac{0.6077}{\eta} - 2\log(0.9069 - \eta) + 0.23\log(\eta) + \log(1.2732\eta) \tag{3.14}$$

For the evaluation of the Helmholtz free energy of the square well solid, Bolhuis and colleagues (Bolhuis, Hagen & Frenkel, 1994) have derived a functional form of N_p/N by fitting their simulation data into a model.

$$\frac{N_p(\delta, \varepsilon, x)}{N} = \frac{3}{2} + \frac{3}{2} erf\left(\frac{3}{4}x - \frac{1}{2}\right) + e^{-x} \sum_{i,j,k=0}^{1,2,6} c_{ijk} \delta^i \varepsilon^{-j} x^k \tag{3.15}$$

where c_{ijk} are the fitted coefficients, and the explicit form of x is exhibited below.

$$x = \frac{\delta}{\left(\frac{\rho_0}{\rho}\right)^{0.5} - 1} \tag{3.16}$$

Although this equation has been derived for the pure square well solid, in the current study, it has been used to calculate the free energy of a binary mixture consisted of hard disk and square well solid particles under the assumption of a pseudo-single-component model, as mentioned before. For that purpose, coefficients for the Monte Carlo simulation data fit for our results of the binary mixture are obtained by using the nonlinear multidimensional regression function provided by MATLAB (MathWorks, Natick, MA, USA). Before evaluating the new coefficients for the binary mixture case, regression analysis has also been performed on the pure square well Monte Carlo dataset, which has been calculated for the current study. A set of new regression coefficients are found (Table 3.1), and the results of the calculation of Eq 3.15 with the new coefficients are compared with the results obtained via the original coefficients to validate the accuracy of the regression analysis. The original coefficient can be found in Table 1 of the reference study (Bolhuis, Hagen & Frenkel, 1994). The model equation results calculated via the regression coefficients, and the simulation data are drawn in the same figure to show the accuracy of the model equation (Fig 3.5).

Table 3.2. Fitting coefficients for the model that has been computed for the pure square well data.

i	***j***	***K***						
		0	1	2	3	4	5	6
0	0	2.2714	-1.4632	0.9594	-4.0919	3.4633	-2.6562	-1.0668
0	1	0.2262	3.1229	-0.035	-0.5273	1.6103	0.0002	-0.2105
0	2	7.3919	-2.0908	0.9188	-3.2121	1.3188	-0.7939	-0.5908
1	0	-69.7481	9.6954	70.6136	-20.2517	-11.3612	42.9310	29.7028
1	1	3.1915	-46.6812	-21.1724	-16.6389	-1.1455	-12.1479	6.0917
1	2	-178.1813	9.2223	71.7349	-1.2048	-4.6107	17.4717	14.3207

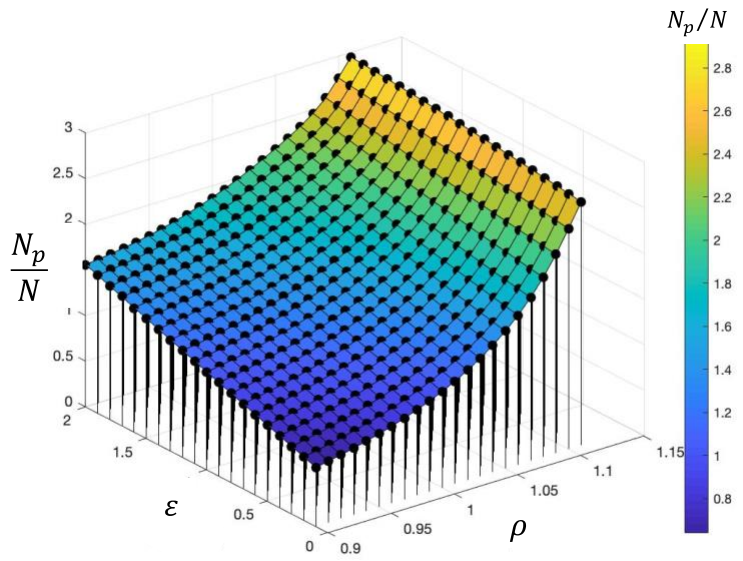

Figure 3.5. Fitting function results (colored surface) and the Monte Carlo simulation data (black dots) for pure square well solid.

The fit successfully matched with the simulation data, with an R^2 value of 0.982. Since the accuracy of the new coefficients has been successfully proved with the pure square well case, the same approach is employed for the binary mixture of hard disk and square well solids. The corresponding coefficients are given in Table 3.2. As can be seen from Figure 3.6, the results of the model equation (colored surface) and the simulation data (black dots) are in a reasonably good agreement, with a model R^2 of 0.896.

Table 3.3. Fitting coefficients for the model that has been computed for the binary mixture of hard disk and square well data

i	***j***	***K***						
		0	1	2	3	4	5	6
0	0	0.50959	3.98785	-36.33678	55.13986	-52.95864	17.92443	-1.32665
0	1	8.39415	-67.27926	207.60521	-308.14618	235.57835	-89.70096	14.12087
0	2	-7.10379	101.32649	-208.02093	307.24666	-241.94340	96.93078	-4.80585
1	0	-53.92652	135.15834	11.27487	6.20017	190.68349	2.23080	-36.55148
1	1	-5.56736	-2.90760	22.96883	-14.98402	16.18487	-5.12853	-16.90736
1	2	-3.761001	-937.15796	130.83027	32.73360	-10.43285	-60.63613	-240.84331

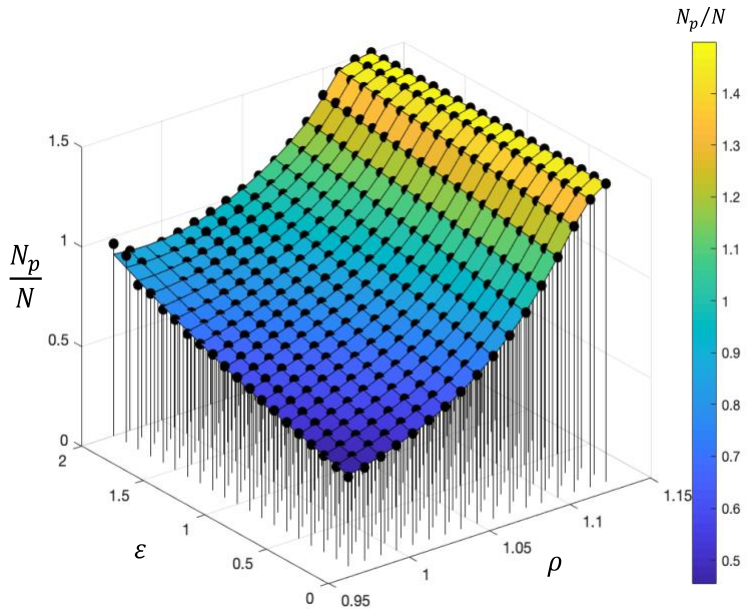

Figure 3.6. Fitting function results (colored surface) and the Monte Carlo simulation data (black dots) for binary mixture of hard disk and square well solid.

To construct the phase diagram of pure square well solids, and the hard disk and square well binary mixtures, the values of the thermodynamic variables at the coexistence should be evaluated first. At the coexistence of two different phases, the chemical potential and pressure values of corresponding phases are required to be identical. Once free energies are evaluated as described in the previous section, coexistence pressure P_{coex} can be calculated, as shown in Eq. 3.20 considering the thermodynamic relations given between Equations 3.17 and 3.19.

$$F = U - TS + \mu N \tag{3.17}$$

The differential form of Equation 3.17 can be written as follows,

$$dF = -SdT - PdV + \mu dN \tag{3.18}$$

where μ is the chemical potential, and S is the entropy. Hence, the pressure is computed, as shown in Equation 3.19.

$$\left(\frac{\partial F}{\partial V}\right)_{T,N} = -P \tag{3.19}$$

Therefore, at the equilibrium Eq. 3.20 should be satisfied.

$$\left.\frac{\partial F}{\partial V}\right|_{V=V1} = \left.\frac{\partial F}{\partial V}\right|_{V=V2} = -P_{coex} \tag{3.20}$$

For two phases being at coexistence, in addition to having the same pressure value, they should have the same chemical potential as well. The thermodynamic relations to calculate the chemical potential for such conditions are summarized below.

$$G = U - TS + pV + \mu N \tag{3.21}$$

By substituting the *F* term in Eq. 3.21 with 3.17, Eq. 3.22 is obtained.

$$G = F + pV + \mu N \tag{3.22}$$

The differential version is,

$$dG = -SdT + VdP + \mu dN \tag{3.23}$$

Hence, the chemical potential is calculated as,

$$\left(\frac{\partial G}{\partial N}\right)_{T,P} = \mu \tag{3.24}$$

For a constant number of particles, with the combination of Eq. 3.21 and Eq. 3.25 chemical potential is calculated as shown in Eq. 3.26.

$$U = TS - PV + \mu N \tag{3.25}$$

$$\mu = \frac{G}{N} \tag{3.26}$$

For the purpose of computing the phase coexistence densities and at a fixed temperature, double tangent construction is applied on Helmholtz free energy vs. volume plot. This method is conceptually identical to Maxwell's equal-area rule, which is applied for a system with measured Gibbs free energy. Since Helmholtz free energy values were computed for the systems in the current study, instead of Maxwell's equal-area rule, double tangent construction is implemented.

For better visualization of the method, an example diagram of free energy vs. volume, which was obtained for the binary mixture of 75% SW - 25% HD case at T = 0.68, is given in Figure 3.7. Briefly, two minima, which have the same pressure on the free energy curve, namely two points on the free energy curve with the same derivatives, are detected as A (F_1, V_1) and B (F_2, V_2). These points indicate the coexistence densities of two different states for a given temperature. These two points on the line also satisfy the necessary conditions to possess the same value of Gibbs free energy. If two points on the curve of $\left(\frac{\partial F}{\partial V}\right)_{T,N}$ are linked with a line, and the slope of this line is also the negative of the pressure value at these particular points, the requirement for having the same Gibbs free energy, hence the chemical potential values are met. The equation of the straight-line which binds points A and B can be written as,

$$y_{DTC} = F_1 + \left(\frac{F_2 - F_1}{V_2 - V_1}\right)(V - V_1) \tag{3.27}$$

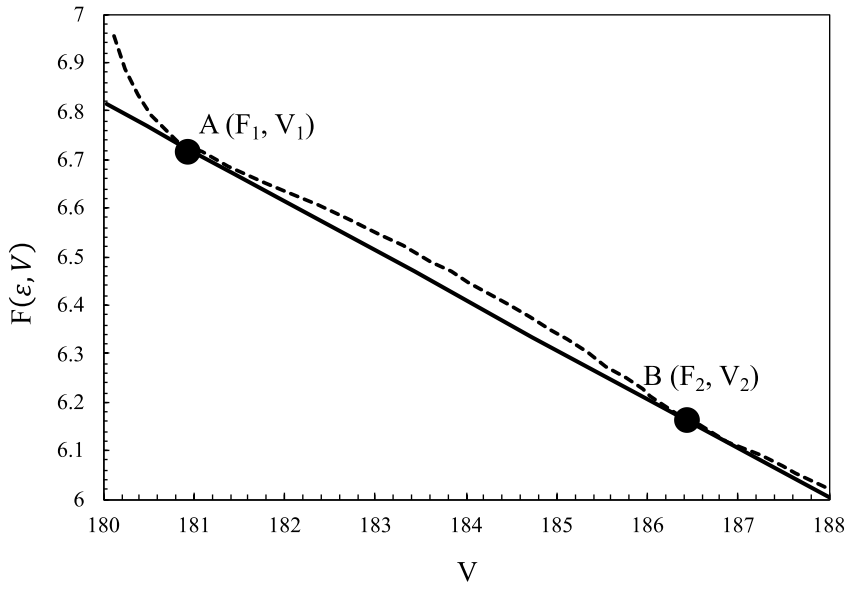

Figure 3.7. Representation of the identification of the two coexistence volumes for the binary mixture of 75% SW - 25% HD by the double tangent construction at T = 0.68.

3.4 Results and Discussion

In this section, the results, including phase diagrams and phase transition behavior for the pure square well and the binary mixture of hard disk and square well, are presented. Firstly, the current study results that are obtained for the pure square well solid have been validated. Following this, the effect of adding hard disk components into pure square well systems on the transition characteristics are discussed.

Firstly, the effect of the composition mixture of the particles on the potential energy with the changing inverse temperature (ε) was investigated. Although the study mainly focuses on the results of the 75% square well and 25% hard disk particle mixtures, the effect of additional hard disk components with different compositions on the potential energy of the system has been examined (Figure 3.8).

For the sake of a clearer exhibition for the potential energy, the term $<U>\varepsilon$ instead of where $<U>$ values are directly taken from the simulation output dataset, for a specific range of inverse temperature, ε. For a constant ε, the pure square well particle system has the lowest average potential energy, which is an expected result, since the square well potential has the potential energy of $-\varepsilon$ in its natural form (Figure 3.2). Hence, as the impurity caused by hard disk components increases, the average potential energy also increases. This representation also serves as proof of the fact that the simulation code modified explicitly for this study is capable of generating physically accurate datasets. It is also noteworthy to mention that, for the maximum hard disk component concentration case, the values of $<U>\varepsilon$ are actually in the order of -10^{-6}, although the results seem "0" from the plot, due to the lack of the step values that y-axis possesses.

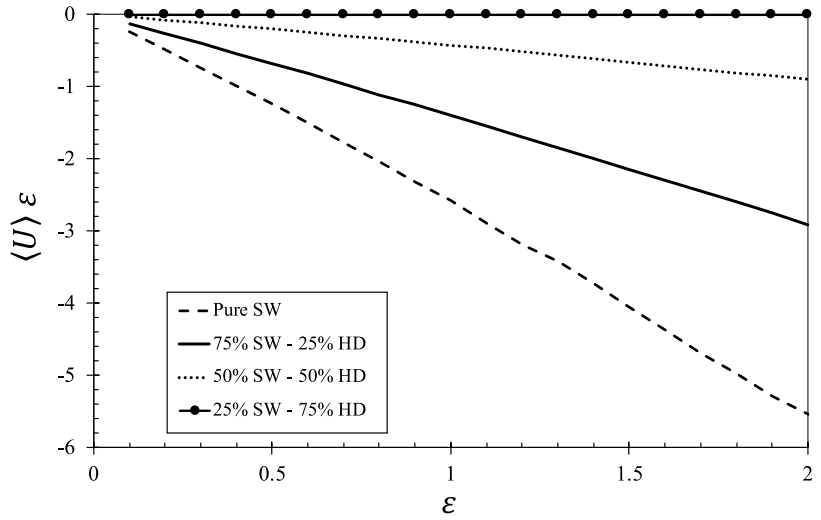

Figure 3.8. Effect of change in the square well compositions on the average potential energy as a function of ε.

As the starting point, free energy values have been computed for the pure square well solids at the range of attraction (δ) value of 0.04. With that respect, the phase diagram of the pure square well solid has been constructed and represented with the black diamonds in Figure 3.8. Another main objective of obtaining the mentioned square well phase diagram is to validate the computation steps which have been used throughout the free energy calculations and the double tangent construction method that has been employed for building the phase diagram. As can be seen from Figure 3.9, the black diamond points, which represent the current study results, match almost perfectly with the results of the reference study, which is shown in Figure 3.3 (Bolhuis, Hagen & Frenkel, 1994). An accurate calculation of the coexistence densities was achieved. The isostructural solid-solid phase transition is visually supported with the snapshots from two different thermodynamic densities at points A and B having the same symmetric structure. An unpaired t-test was applied to understand the significance of the change between the two datasets. By conventional criteria, this difference is considered to be not statistically significant, with a *p*-value of 0.9765 (GraphPad, San Diego, CA, USA).

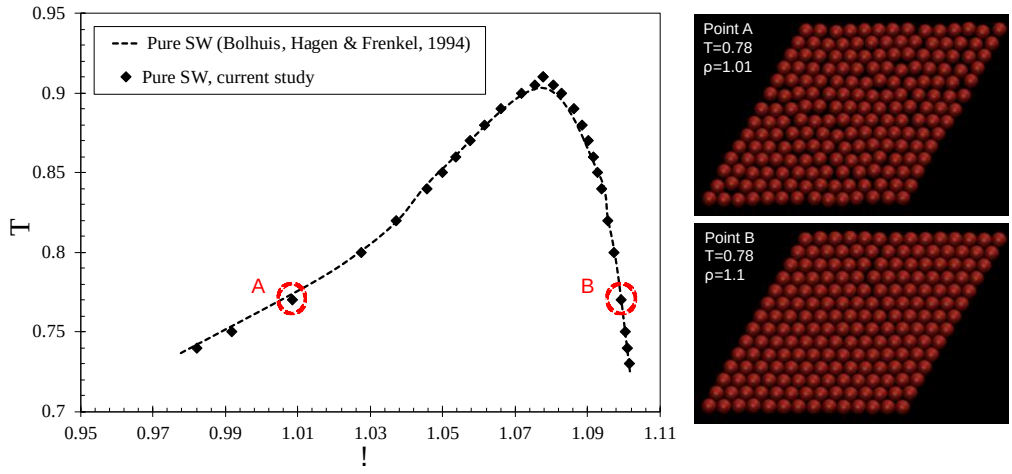

Figure 3.9. Phase diagram of the pure square well solid with δ = 0.04. Black diamond points are the current study results, and the dashed line represents the reference study results (Bolhuis, Hagen & Frenkel, 1994)

The aid of Figure 3.10 can explain the effect of the addition of hard disk solid particles into the pure square well solid on the phase transition. As can be extracted from the figure, for both pure square well (SW) and square well and hard disk (SW-HD) binary mixture systems, solid phases have been divided into two branches (S_1 and S_2) which are distinguished by different densities. Both cases were observed to be isostructural solid-solid phase transitions since for both systems; the transitions are stopped at corresponding critical temperatures (T_C). The reason behind the T_C for the binary mixture is lower than the pure SW system can be explained as; HD components exhibit higher potential energies than SW do; hence, the phase transition could be triggered at a lower temperature than a pure SW system. The slight decrease in the critical temperature can be attributed to the decreased intensity of the intermolecular forces with the addition of HD since they promote the presence of looser contacts between two particles, resulting a decrease in the mixture case. It was also observed that a tighter range of density difference between two branches of the binary mixture case was present when compared to the pure SW case. The studies performed by Bolhuis and co-workers (Bolhuis, Frenkel, and Waals 1994) and Tejero et al. (1994) presented the decrease in the value of δ causes a reduction in the density gap for pure

square well solids. Although parameter δ was kept constant for the current study, this output can also be adapted into our system, where adding a portion of HD particles diminishes the attractive range as if the delta was decreased in the overall system. This adjustability of the density gap between two solid phases can further be examined by changing the concentration of contaminant particles (in our case, HD) to be added to the system. Thus, the design of the desired materials can be further fine-tuned with the expandability of the final solid structure. Configurational snapshots that were prepared with the visual molecular dynamics (VMD) software (UIUC, IL, USA) were also given for two distinct points at S_1 and S_2 branches of the 75% SW – 25% HD phase transition curve. As seen in the snapshots, the symmetry of the lattice structures was preserved, which is another indication of the system going through an isostructural solid-solid phase transition.

As it was briefly summarized in the Section 3.2, there are various rigorous theoretical and numerical methods to compute the free energy. Due to the fact that pure square well free energy calculation procedure has been followed with the modifications for the binary mixture in this study, one of alternative methods to compute the free energy which are originated for the binary mixtures will be explained briefly. Typically, when a phase diagram is trying to be constructed for a binary mixture, simulations of two different phases at the same T and P are individually performed by using the Gibbs ensemble. Then, a method called particle exchange is performed between two individually exploited phases to characterize the phase equilibria behavior of the system as a whole. However, the particle exchange method is challenging when one of the phases is solid, due to the fact that the density is very highl and a large number of particle transfer attempts are needed in between the corresponding phases. Therefore, Kofke (1993) firstly proposed the Gibbs-Duhem integration instead of Gibbs ensemble, to be used for the cases where a solid phase is involved in the system. The difference between two previously mentioned methods is replacing challenging

particle exchange step with the numerical integration of the Clasius-Clapeyron equation. Basically, this scope of the implementation of this equation is to determine the behavior of the field variables, which are kept constant in between the coexisting phases but vary throughout the phase equilibria line. To execute this integration, an initial condition of one component with the parameters including its temperature, pure component fugacity, and the composition are needed. A brief assumption is also made during this integration, where one of the species is assumed to be dilute, to be able to obey Henry's Law.

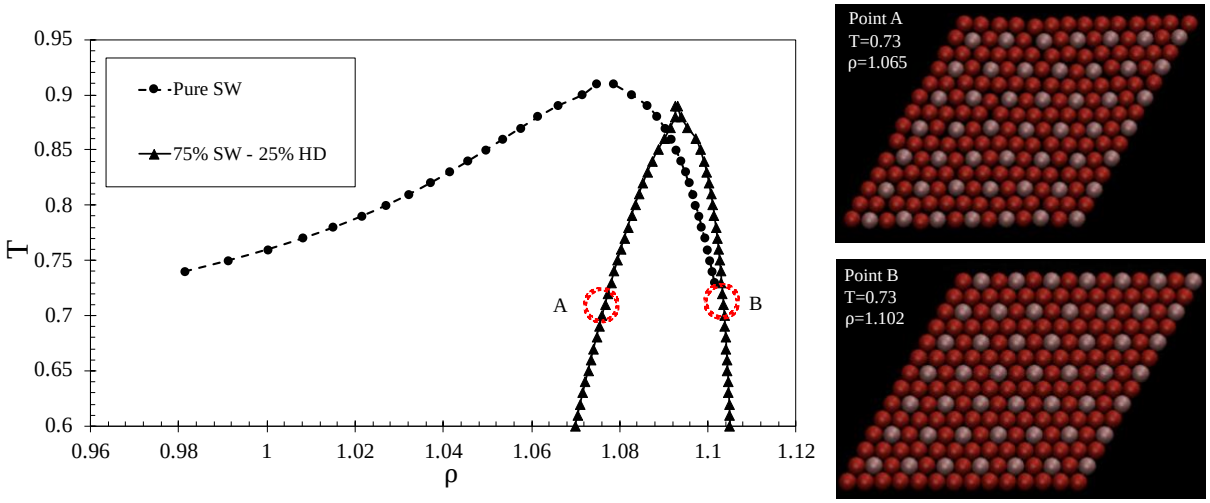

Figure 3.10. Isostructural solid-solid phase transition behaviors for pure square well (SW) and 75% SW - 25% hard disk (HD) binary mixture.

For our study, fcc pure square well coexistence data can be used as the initial condition. The output of the Gibbs-Duhem integration directly gives the temperature values as a function of fugacity, and with these results, the corresponding T-x diagrams can be created for the solid-solid and/or solid-liquid phases.

3.5 Conclusion

In this study, the phase behavior of the binary mixtures consisted of hard disk and square well solid particles has been examined. For that purpose, average potential energy data has been

generated by using canonical ensemble Monte Carlo simulations for 196 particles. After the free energy calculations with the assumption of pseudo-single-component approach, phase diagrams of the pure square well and 75% square well and 25% hard disk binary mixture have been investigated for a very short range of attraction as 0.04. The results showed that the isostructural phase transition, which was already exhibited for pure square well solid particles, occurs in the presence of the additional hard disk particles as well. The existence of the critical temperature for solid-solid transition is the indication of the isostructural phase transition. Additionally, configurational snapshots have been provided as a visual representation, to indicate the presence of conservation in the configurational symmetry at two different solid phases. Critical values of temperature and density were found as 0.87 & 1.09 for the solid-solid isostructural phase transition. In the following steps of this study, as the future work, current results can be validated with alternative methods and the study could be enhanced with the introduction of new parameter effects including varying the diameter ratio and compositions with the implementation of Gibbs-Duhem integration method.

REFERENCES

Alder, B. J., W. E. Alley, and M. Rigby. 1974. “Corrections to the Van Der Waals Model for Mixtures and for the Diffusion Coefficient.” *Physica* 73 (1): 143–55.

Alder, B. J., W. G. Hoover, and D. A. Young. 1968a. “Studies in Molecular Dynamics. V. High-Density Equation of State and Entropy for Hard Disks and Spheres.” *The Journal of Chemical Physics* 49 (8): 3688–96.

Alder, B. J., and T. E. Wainwright. 1962. “Phase Transition in Elastic Disks.” *Physical Review* 127 (2): 359–61.

Allen, Michael P., and Dominic J. Tildesley. *Computer simulation of liquids*. Oxford university press, 2017.

Armas-Pérez, Julio C., Jacqueline Quintana-H, Gustavo A. Chapela, Enrique Velasco, and Guillermo Navascués. 2014. “Phase Diagram of a Square well Model in Two Dimensions.” *Journal of Chemical Physics* 140 (6): 1–8.

Baletto, Francesca, and Riccardo Ferrando. 2005. “Structural Properties of Nanoclusters: Energetic, Thermodynamic, and Kinetic Effects.” *Reviews of Modern Physics* 77 (1): 371–423.

Beck, Thomas L, R Stephen Berry, Thomas L Becka, and R Stephen Berry. 1988. “The Interplay of Structure and Dynamics in the Melting of Small Clusters.” *The Journal of Chemical Physics* 88: 4473.

Beck, Thomas L., David M. Leitner, and R. Stephen Berry. "Melting and phase space transitions in small clusters: Spectral characteristics, dimensions, and K entropy." *The Journal of chemical physics* 89, no. 3 (1988): 1681-1694.

Beltran-Villegas, D.J., Bevan, M.A., 2011. Free energy landscapes for colloidal crystal assembly. Soft Matter 7, 3280–3285. 395

Beltran-Villegas, D.J., Sehgal, R.M., Maroudas, D., Ford, D.M., Bevan, M.A., 2011. A smoluchowski model of crystallization dynamics of small colloidal clusters. The Journal of chemical physics 135, 154506

Beltran-Villegas, J, D., Sehgal, R.M., Maroudas, D., Ford, D.M., Bevan, M.A., 2012. Colloidal cluster crystallization dynamics. The Journal of chemical physics 137, 134901.

Bernard, Etienne P., and Werner Krauth. 2011. "Two-Step Melting in Two Dimensions: First-Order Liquid-Hexatic Transition." *Physical Review Letters* 107 (15).

Berry, R. Stephen. 2002. "The Amazing Phases of Small Systems." *Comptes Rendus Physique* 3 (3): 319–26.

Berry, R S, and Boris M Smirnov. 2009. "Phase Transitions in Various Kinds of Clusters." *Physics-Uspekhi* 52 (2): 137–64.

Berry, R. Stephen, Thomas L. Beck, Heidi L. Davis, and J. Jellinek. "Solid-liquid phase behavior in micro clusters." *Evolution of Size Effects in Chemical Dynamics, Part* 2, no. 2 (1988): 1.

Berry, R Stephen, Julius Jellinek, and Grigory Natanson. 1984. "Physical Review A Volume 30, Number 2 Melting Of Clusters And Melting."

Bevan, Michael A., David M. Ford, Martha A. Grover, Benjamin Shapiro, Dimitrios Maroudas, Yuguang Yang, Raghuram Thyagarajan, Xun Tang, and Ray M. Sehgal. 2015. "Controlling Assembly of Colloidal Particles into Structured Objects: Basic Strategy and a Case Study." *Journal of Process Control* 27: 64–75.

Bokis, C. P., and M. D. Donohue. 1995. "A New Equation of State for Square well Molecules." *International Journal of Thermophysics* 16 (5): 1277–86.

Bolhuis, Peter, and Daan Frenkel. "Prediction of an expanded-to-condensed transition in colloidal crystals." *Physical review letters* 72, no. 14 (1994): 2211.

Bolhuis, Peter, Maarten Hagen, and Daan Frenkel. 1994. "Isostructural Solid-Solid Transition in Crystalline Systems with Short-Ranged Interaction." *Physical Review E* 50 (6): 4880–90.

Borrmann, Peter, Oliver Mülken, and Jens Harting. 2000. “Classification of Phase Transitions in Small Systems.” *Physical Review Letters* 84 (16): 3511–14.

Brilliantov, Nikolai V, and John P Valleau. 1998. “Thermodynamic Scaling Monte Carlo Study of the Liquid–Gas Transition in the Square–Well Fluid.” *The Journal of Chemical Physics* 108: 174505.

Calvo, F., P. Parneix, and M. Basire. 2009. “Quantum Densities of States of Fluxional Polyatomic Systems from a Superposition Approximation.” *Journal of Chemical Physics* 130 (15).

Chang, Jaeeon, and Stanley I. Sandler. "A completely analytic perturbation theory for the square well fluid of variable well width." *Molecular Physics* 81, no. 3 (1994): 745-765.

Coifman, R. R., I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler. 2008. “Diffusion Maps, Reduction Coordinates, and Low Dimensional Representation of Stochastic Systems.” *Multiscale Modeling & Simulation* 7 (2): 842–64.

Coifman, R. R., S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. 2005. “Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Multiscale Methods.” *Proceedings of the National Academy of Sciences* 102 (21): 7432–37.

Coifman, Ronald R., Stephane Lafon, Ann B. Lee, Mauro Maggioni, Boaz Nadler, Frederick Warner, and Steven W. Zucker. "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps." *Proceedings of the national academy of sciences* 102, no. 21 (2005): 7426-7431.

Coifman, Ronald R., and Stéphane Lafon. 2006. “Diffusion Maps.” *Applied and Computational Harmonic Analysis* 21 (1): 5–30.

Das, Ritankar, and David J. Wales. "Energy landscapes for a machine-learning prediction of patient discharge." *Physical Review E* 93, no. 6 (2016): 063310.

Davis, Heidi L., Julius Jellinek, and R. Stephen Berry. 1987. “Melting and Freezing in Isothermal Ar13 Clusters.” *The Journal of Chemical Physics* 86 (11): 6456–64.

Denton', A R, and N W Ashcroft. 1990. "Weighted-Density-Functional Theory of Nonuniform Fluid Mixtures: Application to Freezing of Binary Hard-Sphere Mixtures." *Physical Review A*. Vol. 42.

Dijkstra, Marjolein. 2002. "Phase Behavior of Hard Spheres with a Short-Range Yukawa Attraction." *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 66 (2): 1–5.

Dijkstra, Marjolein. "Modelling and Simulation.", 2014.

Dsilva, Carmeline J., Ronen Talmon, C. William Gear, Ronald R. Coifman, and Ioannis G. Kevrekidis. 2015. "Data-Driven Reduction for Multiscale Stochastic Dynamical Systems" 15 (3): 1327–51.

Eldridge, M. D., P. A. Madden, and D. Frenkel. 1993. "The Stability of the AB13 Crystal in a Binary Hard Sphere System." *Molecular Physics* 79 (1): 105–20.

Elliott, J. Richard, and Liegi Hu. 1999. "Vapor-Liquid Equilibria of Square well Spheres." *The Journal of Chemical Physics* 110 (6): 3043–48.

Erpenbeck, Jerome J., and Marshall Luban. 1985. "Equation of State of the Classical Hard disk Fluid." *Physical Review A* 32 (5): 2920–22.

Ferguson, A. L., A. Z. Panagiotopoulos, P. G. Debenedetti, and I. G. Kevrekidis. 2010. "Systematic Determination of Order Parameters for Chain Dynamics Using Diffusion Maps." *Proceedings of the National Academy of Sciences* 107 (31): 13597–602.

Ferguson, Andrew L., Athanassios Z. Panagiotopoulos, Pablo G. Debenedetti, and Ioannis G. Kevrekidis. 2011. "Integrating Diffusion Maps with Umbrella Sampling: Application to Alanine Dipeptide." *Journal of Chemical Physics* 134 (13).

Ferguson, Andrew L., Athanassios Z. Panagiotopoulos, Ioannis G. Kevrekidis, and Pablo G. Debenedetti. 2011. "Nonlinear Dimensionality Reduction in Molecular Simulation: The Diffusion Map Approach." *Journal of Chemical Physics* 509 (1–3): 1–11.

Fernández, Ángela, Ana M. González, Julia Díaz, and José R. Dorronsoro. "Diffusion maps for dimensionality reduction and visualization of meteorological data." *Neurocomputing* 163 (2015): 25-37.

Foffi, Giuseppe, Gavin D. McCullagh, Aonghus Lawlor, Emanuela Zaccarelli, Kenneth A. Dawson, Francesco Sciortino, Piero Tartaglia, Davide Pini, and George Stell. 2002. "Phase Equilibria and Glass Transition in Colloidal Systems with Short-Ranged Attractive Interactions: Application to Protein Crystallization." *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 65 (3): 1–17.

Ford, David M., Aditya Dendukuri, Gulce Kalyoncu, Khoa Luu, and Matthew J. Patitz. "Machine Learning to Identify Variables in Thermodynamically Small Systems." *Computers & Chemical Engineering* (2020): 106989.

Frenkel, Daan. "Free-energy calculations." *NATO ASI Series. Series E: Applied sciences* 205 (1991): 85-117.

Galdino, Ramon V., Clayton A. Benevides, and Rômulo P. Tenório. "Diffusion maps of Bacillus subtilis biofilms via magnetic resonance imaging highlight a complex network of channels." *Colloids and Surfaces B: Biointerfaces* (2020): 110905.

García, Álvaro González, Joeri Opdam, Remco Tuinier, and Mark Vis. 2018. "Isostructural Solid–Solid Coexistence of Colloid–Polymer Mixtures." *Chemical Physics Letters* 709 (July): 16–20.

Giacometti, Achille, Fred Lado, Julio Largo, Giorgio Pastore, and Francesco Sciortino. 2009. "Phase Diagram and Structural Properties of a Simple Model for One-Patch Particles." *The Journal of Chemical Physics* 131: 144504.

Grigoryan, Valeri G., and Michael Springborg. 2019. "Temperature and Isomeric Effects in Nanoclusters." *Physical Chemistry Chemical Physics* 21 (10): 5646–54.

Gulati, Harpreet S., and Carol K. Hall. "Fluids and fluid mixtures containing square well diatomics: Equations of state and canonical molecular dynamics simulation." *The Journal of chemical physics* 107, no. 10 (1997): 3930-3946.

Gunsteren, W. F. van, and H. J.C. Berendsen. 1987. “Thermodynamic Cycle Integration by Computer Simulation as a Tool for Obtaining Free Energy Differences in Molecular Chemistry.” *Journal of Computer-Aided Molecular Design* 1 (2): 171–76.

Hannon, James B., and Ruud M. Tromp. "Low-energy electron microscopy of surface phase transitions." *Annual Review of Materials Research* 33, no. 1 (2003): 263-288.

Henderson, Douglas. 1974. “Perturbation Theory for a Mixture of Hard Spheres and Square well Molecules.” *The Journal of Chemical Physics* 61 (3): 926–31.

Hill, Terrell L. "NanoSeconds-Perspective: Nanothermodynamics." *Nano Letters* 1, no. 3 (2001): 111-112.

Hill, T. L. Thermodynamics of Some Small Systems. Vol. 1. Benjamin, 1963.

Hill, T. L. Thermodynamics of Small Systems; Benjamin: New York, Part II,1964

Hinde, Robert J, R Stephen Berry, and David J Wales. 1992. “Chaos in Small Clusters of Inert Gas Atoms.” *The Journal of Chemical Physics* 96: 3263.

Hoare, M. R., and P. Pal. "Statistics and stability of small assemblies of atoms." *Journal of Crystal Growth* 17 (1972): 77-96.

Honeycutt, J. Dana, and Hans C. Andersen. 1987. “Molecular Dynamics Study of Melting and Freezing of Small Lennard-Jones Clusters.” *Journal of Physical Chemistry* 91 (19): 4950–63.

Hu, Jiawen, and Yang Xin Yu. 2009. “High-Order Virial Coefficients and Equation of State for Hard Sphere and Hard Disk Systems.” *Physical Chemistry Chemical Physics* 11 (41): 9382–90.

Hub, Jochen S., and Bert L. De Groot. "Detection of functional modes in protein dynamics." *PLoS Comput Biol* 5, no. 8 (2009): e1000480.

Huerta, A., V. Carrasco-Fadanelli, and A. Trokhymchuk. 2012. “Towards Frustration of Freezing Transition in a Binary Hard disk Mixture.” *Condensed Matter Physics* 15 (4).

Huerta, Adrián, Douglas Henderson, and Andrij Trokhymchuk. n.d. “Freezing of Two-Dimensional Hard Disks.”

Jellinek, Julius, Thomas L Beck, R Stephen Berry, and A Stephen Berry. 1986. “Solid–Liquid Phase Changes in Simulated Isoenergetic Ar13.” *The Journal of Chemical Physics* 84: 545.

Jia, Bo, Biting Yu, Qi Wu, Xinshe Yang, Chuanfeng Wei, Rob Law, and Shan Fu. "Hybrid local diffusion maps and improved cuckoo search algorithm for multiclass dataset analysis." *Neurocomputing* 189 (2016): 106-116.

Jortner, Joshua, and C. N.R. Rao. 2002. “Nanostructured Advanced Materials. Perspectives and Directions.” *Pure and Applied Chemistry* 74 (9): 1491–1506.

Kaelberer, J. B., and R. D. Etters. 1976. “Phase Transitions in Small Clusters of Atoms.” *The Journal of Chemical Physics* 66 (7): 3233–39.

Kawamura, H. 2005. “A Simple Theory of Hard Disk Transition.” *Progress of Theoretical Physics* 61 (6): 1584–96.

Kim, Sang Beom, Carmeline J. Dsilva, Ioannis G. Kevrekidis, and Pablo G. Debenedetti. "Systematic characterization of protein folding pathways using diffusion maps: Application to Trp-cage miniprotein." *The Journal of chemical physics* 142, no. 8 (2015): 02B613_1.

Kofke, David A. "Gibbs-Duhem integration: a new method for direct evaluation of phase coexistence by molecular simulation." *Molecular Physics* 78, no. 6 (1993): 1331-1336.

Kranendonk, W G T, and D Frenkel. 1989. “Computer Simulation of Solid-Liquid Coexistence in Binary Hard-Sphere Mixtures.” *J. Phys.: Condens. Matter*. Vol. 1.

Koronaki, E. D., A. M. Nikas, and A. G. Boudouvis. "A data-driven reduced-order model of nonlinear processes based on Diffusion Maps and Artificial Neural Networks." *Chemical Engineering Journal* (2020): 125475.

Kummerfeld, Jonathan K., Toby S. Hudson, and Peter Harrowell. 2008. “The Densest Packing of AB Binary Hard-Sphere Homogeneous Compounds across All Size Ratios.” *Journal of Physical Chemistry B* 112 (35): 10773–76.

Lancewicki, Tomer, and Mayer Aladjem. "Multi-target shrinkage estimation for covariance matrices." *IEEE Transactions on Signal Processing* 62, no. 24 (2014): 6380-6390.

Lerbret, A, P Bordat, F Affouard, M Descamps, and F Migliardo. 2005. “How Homogeneous Are the Trehalose , Maltose , and Sucrose Water Solutions ? An Insight from Molecular Dynamics Simulations,” 11046–57.

Likos, C. N., Zs T. Németh, and H. Lowen. "Density-functional theory of solid-to-solid isostructural transitions." *Journal of Physics: Condensed Matter* 6, no. 50 (1994): 10965.

Lin, Shang-Chun Chun, and Martin Oettel. 2018. “Phase Diagrams and Crystal-Fluid Surface Tensions in Additive and Nonadditive Two-Dimensional Binary Hard disk Mixtures.” *Physical Review E* 98 (1): 1–12.

Long, A.W., Ferguson, A.L., 2014. Nonlinear machine learning of patchy colloid self-assembly pathways and mechanisms. The Journal of Physical Chemistry B 118, 4228–4244.

Love, J. Christopher, Lara A. Estroff, Jennah K. Kriebel, Ralph G. Nuzzo, and George M. Whitesides. 2005. *Self-Assembled Monolayers of Thiolates on Metals as a Form of Nanotechnology*. *Chemical Reviews*. Vol. 105.

Mansbach, Rachael A., and Andrew L. Ferguson. 2015. “Machine Learning of Single Molecule Free Energy Surfaces and the Impact of Chemistry and Environment upon Structure and Dynamics.” *Journal of Chemical Physics* 142 (10).

Marks, L. D., and L. Peng. 2016. “Nanoparticle Shape, Thermodynamics and Kinetics.” *Journal of Physics Condensed Matter* 28 (5).

Méndez-Maldonado, G. Arlette, Gustavo A. Chapela, José Adrián Martínez-González, José Antonio Moreno, Enrique Díaz-Herrera, and José Alejandre. "Fluid-solid coexistence from two-phase simulations: Binary colloidal mixtures and square well systems." *The Journal of chemical physics* 142, no. 5 (2015): 054501.

Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. "Equation of state calculations by fast computing machines." *The journal of chemical physics* 21, no. 6 (1953): 1087-1092.

McBride, Carl, and Carlos Vega. 2002. “Fluid Solid Equilibrium for Two Dimensional Tangent Hard Disk Chains from Wertheim’s Perturbation Theory.” *Journal of Chemical Physics* 116 (5): 1757–59.

Miguel, Rodrigo De, and J. Miguel Rubí. 2017. “Thermodynamics Far from the Thermodynamic Limit.” *Journal of Physical Chemistry B* 121 (45): 10429–34.

Mitus, A. C., H. Weber, and D. Marx. "Local structure analysis of the hard disk fluid near melting." *Physical Review E* 55, no. 6 (1997): 6855.

Mohazzabi, Pirooz, and G. Ali Mansoori. 2018. “Why Nanosystems and Macroscopic Systems Behave Differently” 1 (1): 53–60.

Monson, P. A., and D. A. Kofke. "Solid-fluid equilibrium: Insights from simple molecular models." *Advances in Chemical Physics* 115 (2000): 113-180.

Nadler, Boaz, Stéphane Lafon, Ronald R. Coifman, and Ioannis G. Kevrekidis. 2006. “Diffusion Maps, Spectral Clustering and Reaction Coordinates of Dynamical Systems.” *Applied and Computational Harmonic Analysis* 21 (1): 113–27.

Nedialkova, Lilia V., Miguel A. Amat, Ioannis G. Kevrekidis, and Gerhard Hummer. 2014. “Diffusion Maps, Clustering and Fuzzy Markov Modeling in Peptide Folding Transitions.” *The Journal of Chemical Physics* 141 (11): 114102.

Nieminen, Paavo, Ilkka Pölönen, and Tuomo Sipola. "Research literature clustering using diffusion maps." *Journal of Informetrics* 7, no. 4 (2013): 874-886.

Orkoulas, Gerassimos, and Athanassios Z. Panagiotopoulos. 1999. “Phase Behavior of the Restricted Primitive Model and Square well Fluids from Monte Carlo Simulations in the Grand Canonical Ensemble.” *Journal of Chemical Physics* 110 (3): 1581–90.

Pan, Albert C., and David Chandler. "Dynamics of nucleation in the Ising model." *The Journal of Physical Chemistry B* 108, no. 51 (2004): 19681-19686.

Palomares-Baez, Juan Pedro, Emanuele Panizon, and Riccardo Ferrando. 2017. “Nanoscale Effects on Phase Separation.” *Nano Letters* 17 (9): 5394–5401.

Parker, Greg. "Encyclopedia of materials: science and technology." (2001): 3703-3707.

Philippe, T. 2017. “Nucleation and Superstabilization in Small Systems.” *Physical Review E* 96 (3): 1–6.

Rosenfeld, Y., J. E. Avron, S. Goshen, and R. Thieberger. "Variational approach to solids and the hard sphere melting law." *The Journal of Chemical Physics* 66, no. 6 (1977): 2758-2760.

Qian, Hong. 2012. “Hill’s Small Systems Nanothermodynamics: A Simple Macromolecular Partition Problem with a Statistical Perspective.” *J Biol Phys* 38: 201–7.

Quirke, N., and Ping Sheng. 1984. “The Melting Behavior of Small Clusters of Atoms.” *Chemical Physics Letters* 110 (1): 63–66.

Ree, Francis H., and William G. Hoover. 1964. “Fifth and Sixth Virial Coefficients for Hard Spheres and Hard Disks.” *The Journal of Chemical Physics* 40 (4): 939–50.

Reguera, D., R. K. Bowles, Y. Djikaev, and H. Reiss. 2003. “Phase Transitions in Systems Small Enough to Be Clusters.” *Journal of Chemical Physics* 118 (1): 340–53.

Reinhart, Wesley F., Andrew W. Long, Michael P. Howard, Andrew L. Ferguson, and Athanassios Z. Panagiotopoulos. 2017. “Machine Learning for Autonomous Crystal Structure Identification.” *Soft Matter* 13 (27): 4733–45.

Rey, Marcel, Miguel Ángel Fernández-Rodríguez, Mathias Steinacher, Laura Scheidegger, Karen Geisel, Walter Richtering, Todd M. Squires, and Lucio Isa. 2016. “Isostructural Solid-Solid Phase Transition in Monolayers of Soft Core-Shell Particles at Fluid Interfaces: Structure and

Mechanics.” *Soft Matter* 12 (15): 3545–57.

Rohrdanz, Mary A., Wenwei Zheng, Mauro Maggioni, and Cecilia Clementi. 2011. “Determination of Reaction Coordinates via Locally Scaled Diffusion Map.” *Journal of Chemical Physics* 134 (12).

Russo, John, and Nigel B. Wilding. "Disappearance of the hexatic phase in a binary mixture of hard disks." *Physical review letters* 119, no. 11 (2017): 115702.

Sehgal, Ray M., and Dimitrios Maroudas. 2015. “Equilibrium Shape of Colloidal Crystals.” *Langmuir* 31 (42): 11428–37.

Sehgal, Ray M, Dimitrios Maroudas, and David M Ford. 2014. “Phase Behavior of the 38-Atom Lennard-Jones Cluster.” *J. Chem. Phys* 140: 104312.

Silva, M. A A Da, A. Caliri, and B. J. Mokross. 1987. “Phase Transition in a System of Hard Disks by Monte Carlo Simulation.” *Physical Review Letters* 58 (22): 2312–14.

Smithline, S. J., and A. D.J. J Haymet. 1987. “Density Functional Theory for the Freezing of 1:1 Hard Sphere Mixtures.” *The Journal of Chemical Physics* 86 (11): 6486–94.

Singh, R., Xu, J., Berger, B., 2008. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proceedings of the National Academy of Sciences* 105, 12763–12768

Speedy, Robin J. 1999. “Relaxation in Glassforming Liquids and Amorphous Solids.” *The Journal of Chemical Physics* 110: 80901.

Stamati, Hernán, Cecilia Clementi, and Lydia E. Kavraki. "Application of nonlinear dimensionality reduction to characterize the conformational landscape of small peptides." *Proteins: Structure, Function, and Bioinformatics* 78, no. 2 (2010): 223-235.

Straffin, P.D., 1980. Linear algebra in geography: eigenvectors of networks. *Mathematics Magazine* 53, 269–276.

Stukowski, Alexander. "Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool." *Modelling and Simulation in Materials Science and Engineering* 18, no. 1 (2009): 015012.

Tejero, C. F., A. Daanoun, H. N.W. Lekkerkerker, and M. Baus. 1994. “Phase Diagrams of ‘Simple’ Fluids with Extreme Pair Potentials.” *Physical Review Letters* 73 (5): 752–55.

Tester, Jefferson W, and Michael Modell. 1997. “Thermodynamics and Its Applications.” *Book*.

The Cambridge Cluster Database, D. J. Wales, J. P. K. Doye, A. Dullweber, M. P. Hodges, F. Y. Naumkin F. Calvo, J. Hernández-Rojas and T. F. Middleton, URL http://www-wales.ch.cam.ac.uk/CCD.html

Truskett, Thomas M., Salvatore Torquato, Srikanth Sastry, Pablo G. Debenedetti, and Frank H. Stillinger. "Structural precursor to freezing in the hard disk and hard-sphere systems." *Physical review E* 58, no. 3 (1998): 3083.

Vega, Lourdes, Enrique de Miguel, Luis F. Rull, George Jackson, and Ian A. McLure. "Phase equilibria and critical behavior of square-well fluids of variable width by Gibbs ensemble Monte Carlo simulation." *The Journal of chemical physics* 96, no. 3 (1992): 2296-2305.

Velasco, E., G. Navascués, and L. Mederos. "Phase behavior of binary hard-sphere mixtures from perturbation theory." *Physical Review E* 60, no. 3 (1999): 3158.

Vlugt, Thijs J H. 2009. *Introduction to Molecular Simulation and Statistical Thermodynamics*. *Computer*.

Wai, YauYau, JuHsu Chu, ChunChieh Wang, YuChun Lin, Gigin Lin, YungLiang Wan, and JiunJie Wang. "An integrated diffusion map for the analysis of diffusion properties: a feasibility study in patients with acoustic neuroma." *Academic radiology* 16, no. 4 (2009): 428-434.

Wales, D. J., and R. S. Berry. "Local interpretation of chaotic dynamics in a many-body classical Hamiltonian system (Ar3)." *Journal of Physics B: Atomic, Molecular and Optical Physics* 24, no. 14 (1991): L351.

Wang, Helan, Shan Zhou, Kyle D. Gilroy, Zaisheng Cai, and Younan Xia. 2017. “Icosahedral Nanocrystals of Noble Metals: Synthesis and Applications.” *Nano Today* 15: 121–44.

Wang, Hong, and Wenchuan Wang. "Monte Carlo simulation of ternary square well fluid mixtures." *Fluid phase equilibria* 60, no. 1-2 (1990): 11-36.

Warshavsky, Vadim B., and Xueyu Song. 2008. “Fundamental Measure Density Functional Theory Studies on the Freezing of Binary Hard-Sphere and Lennard-Jones Mixtures.” *Journal of Chemical Physics* 129 (3): 1–8.

Wood, W. W. 1968. “Monte Carlo Calculations for Hard Disks in the Isothermal-Isobaric Ensemble.” *The Journal of Chemical Physics* 48 (1): 415–34.

Wood, W W, and J J Erpenbeck. 1976. “Molecular Dynamics and Monte Carlo Calculations in Statistical Mechanics.” *Annual Review of Physical Chemistry* 27 (1): 319–48.

Xu, Rui, Steven Damelin, Boaz Nadler, and Donald C. Wunsch II. "Clustering of high-dimensional gene expression data with feature filtering methods and diffusion maps." *Artificial intelligence in medicine* 48, no. 2-3 (2010): 91-98.

Xue, Yuzhen, Peter J. Ludovice, Martha A. Grover, Lilia V. Nedialkova, Carmeline J. Dsilva, and Ioannis G. Kevrekidis. 2013. “State Reduction in Molecular Simulations.” *Computers and Chemical Engineering* 51: 102–10.

Yang, Yuguang, Raghuram Thyagarajan, David M. Ford, and Michael A. Bevan. 2016. “Dynamic Colloidal Assembly Pathways via Low Dimensional Models.” *Journal of Chemical Physics* 144 (20).

Young, David A., and Berni J. Alder. 1979. “Studies in Molecular Dynamics. XVII. Phase Diagrams for ‘Step’ Potentials in Two and Three Dimensions.” *The Journal of Chemical Physics* 70 (1): 473–81.