Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Physics

Amund Bertheussen

Relaxation process of Yukawa glassy system

Master's thesis in Applied Physics and Mathematics
Supervisor: Raffaela Cabriolu
Co-supervisor: Erika Eiser
June 2022

NTNU
Norwegian University of
Science and Technology

Amund Bertheussen

Relaxation process of Yukawa glassy system

Master's thesis in Applied Physics and Mathematics
Supervisor: Raffaela Cabriolu
Co-supervisor: Erika Eiser
June 2022

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Physics

NTNU
Norwegian University of
Science and Technology

Abstract

The Relaxation-process of glassy systems is of great interest in pharmaceutical, cosmetics, food, and engineering construction fields. As a general feature, when a glass is forming, the viscosity is changing and the dynamics of the system is slowing down greatly. In this Master's thesis, the relaxation process of a Yukawa binary system that is forming glass was investigated. The relaxation time of the glassy sample was measured using common correlation functions and the time-evolution of the typical heterogeneities was studied using a Molecular Dynamics simulation technique. Additionally, a popular machine learning tool was used to quantify the heterogeneity in the samples.

The results showed similar dynamics to related works and all the different regimes known to occur for glassy states were characterized. Furthermore, the machine-learning approach of k-means showed to agree with the heterogeneity analysis done with traditional methods.

Sammendrag

Relaksasjonsprosessen til et glass-system er av høy interesse i mange industrier, blant annet farmasi, kosmetikk, mat og bygge bransjen. Når et glass formes øker som regel viskositeten og dynamikken til systemet blir mye tregere. I denne master oppgaven ble relaksasjon av et Yukawa binært kolloidalt system som danner et glass, undersøkt. Relaksasjonstiden av glass systemet ble målt med korrelasjons funksjoner og tidsutviklingen av typisk heterogenitet ble studert ved hjelp av Molekylær Dynamikk. I tillegg ble en populær maskinlæringsalgoritme brukt til å kvantifisere heterogeniteten i systemene.

Resultatene viste seg å stemme med tidligere relatert arbeid og alle de forskjellige regimene som er er knyttet til glass systemer ble karakterisert. Videre, viste k-means maskinlærings teknikken å stemme med tradisjonelle måter for å analysere den dynamiske heterogeniteten i glass.

Preface

This thesis is submitted as part of the course TFY4900 - Physics, Master's Thesis (30 ECTS) at NTNU[1]. The thesis marks the end of a five-year master's degree in Applied Physics and Mathematics at NTNU, with Applied Physics as the main profile. The work presented in this thesis has been carried out through the first half of 2022 and is partly a continuation of a specialization project conducted during the autumn of 2021. The whole work originates to Dr. Raffaela Cabriolu who has been my supervisor during this period.

The five last years have been a chapter of my life full of hard work and valuable lessons. Not only have my study program given me a unique expertise in physics and mathematics, but it has also offered me general knowledge in other important fields like data science, economics and leadership, all of which I will benefit from later in life. University life and all things associated has given me experiences and friendships I will forever be grateful for. Even though challenges, hard exams and a global pandemic have given me frustration and long nights, I will look back at these years with appreciation.

I would like to thank my friends and family for the support I have experienced along the road to this master's degree. A special thanks will also be expressed to my supervisor Raffaela Cabriolu. I honestly do not think this thesis would be any good without her help and incredible expertise in the field of soft condensed matter physics and molecular dynamics. Even with a packed schedule, she has given me answers and guidance that have made it possible for me to write code for and understand all the different aspects that have led to this thesis.

Finally, acknowledge is sent to PRACE for awarding me and Cabriolu access to Piz Daint, at the Swiss National Supercomputing Centre (CSCS), project number "pr128". This access has given me an incredible experience of utilizing High Performance Clusters and made it possible to simulate the systems required to get accurate measurements used in this thesis.

Trondheim, June 2022
Amund Bertheussen

[1] Norwegian University of Science and Technology

Contents

Chapter 1

Introduction

The specific state of a material, solid, liquid or gas phase, is related to the amount of internal energy in the system. Depending on the phase, the behavior of the material is very different. The solid state is defined as the state with lower energy with respect to the other two states. In this phase, the constituents of the material, i.e. molecules or atoms, are bounded to equilibrium positions by interaction forces in a highly ordered configuration characterized by fixed volume and shape. However, there exists temperatures different than zero, for which the internal energy of the constituents is bigger than the interaction forces such that the constituents can vibrate around their equilibrium points. This is why solids surrounding us every day, are rigid and mostly non-compressible. Increasing the temperature of a solid, determines an increase of the internal energy and the vibrations grow, forcing the solid to expand [1].

By increasing the internal energy of the system even further, the energy of some particles will eventually be sufficient to break the particle free from the interaction forces, determining the loss of the solid order. This breakdown denotes the phase transition from solid to liquid, and the particles are now able to flow. Only short-range order is present in the system, but, unlike a gaseous state, this system is more or less incompressible for most cases. The particle's ability to move around give rise to a phenomenon called diffusion. For an equilibrated state, this diffusion is only driven by random motion of the particles and thereby called self-diffusion. This property of liquids is one of the measurements that could be used to distinguish the liquid from the solid state, tho which no diffusion of particles is present.

An interesting phenomena occurs if the liquefied system undergoes a rapid cooldown, also known as quenching. Because of the short timespan of the quenching process, the energy suddenly decrease and most of the particles will not have 'sufficient time' to assume the order of a low energy state typical of the solid, i.e. no crystallization of the sample. A rapid cooling of a liquid without the presence of crystallization is often called a vitrification process. This results in a system with only short-range ordering of the particles. However, because the particles have low internal energy, they are locked into this almost frozen disordered configur-

ation characterized by a very limited motion without diffusion. This state shares properties with both solid and liquid materials and is known as glass [2]. The glass structure is often referred as an amorphous solid structure. Previously, this state was also called a supercooled liquid because its structure being so similar to a liquid state. [2, 3].

Glass formation through a sudden change in temperature, as introduced here, is one of the most common ways to prepare a glass. However, it is important to specify that beside the quenching method, it is possible to create systems with typical glass behavior by increasing the pressure or density of the system [4]. In this computational study, a glassy state will only be referring to the the one obtained by the process of cooling down the system. Figure 1.1 illustrates the progression of a thought experiment that takes an exemplar material from a solid to a liquid, and, finally to a glassy state reached by decreasing the Temperature.

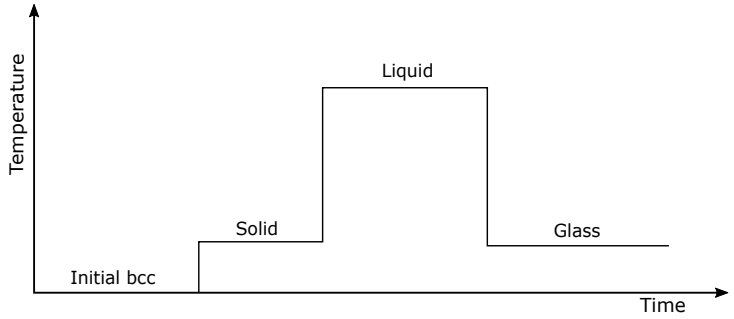


Figure 1.1: Timeline for preparing a glass structure through the process of quenching.

Natural glasses, e.g. obsidian and amber, have existed long before the emergence of life on earth. Synthetic oxide glasses were discovered circa 6,000 years ago and are now ubiquitous and essential for our lives and many technology applications. Despite the importance and the enormous implication in our life and most scientific fields, there is still no consensus on which physical mechanisms underlie the process of glass formation, and, associated to it, there are many unanswered questions that pose challenges and attract the attention of numerous studies [3–5].

The vitrification process is unlike any thermodynamic phase transition, since the material exhibits a spectacular growth of viscosity upon cooling or compression, but only minute structural changes. The so-called supercooled regime– is characterized by viscosity η so large that the material stops flowing on any practical time scale. A somehow arbitrary definition a glass is such that it is characterized by a viscosity value over $10^{12} Pa \cdot s$ or a structural relaxation time τ exceeding 100 s [2, 3]. The gradual rearrangement of the structure towards, ideally, the equilibrium represents the structural relaxation. The most detailed information of the vitrification process is contained in the microscopic relaxation dynamics. Furthermore, what makes this description ideal, is that it can be accessed by both, molecular simulations and experiments [4, 6, 7].

Strictly connected to the structural relaxation is another phenomena that makes glasses very different from ordinary liquids, this is the appearance of Dynamic Heterogeneity (DH). In glasses, the structural relaxation does not take place uniformly across the sample, but rather in clusters, domains or local regions that collectively rearrange while the rest of the super-cooled liquid pertain in a temporary frozen condition [6, 8].

The raising of the viscosity, relaxation times and the appearance of dynamical heterogeneity are universal hallmark features for all glass-forming liquids. They indicate general dynamical slow down around the glass transition of the material. However, how those dynamical hallmark are changing in respect to the external parameters, like temperature and pressure is not universal for all materials and the physical reasons is not fully understood [4, 6, 9, 10].

Scientist of glassy materials are actively searching for a origin that governs this dynamical slowing down in the supercooled liquid, which occurs without any discernible change in structure [10]. Extensive experimental and specially modeling research concluded that there should exist growing dynamic and static correlation lengths associated respectively to the dynamical heterogeneity and to the raise of the viscosity [9, 11, 12]. In particular, the concept of static length scale is associate with the spatial "amorphous" order in glasses and it is considered a characteristic static property that can be measured by equilibrium properties of the material. The concept is based on the Random First Order Transition theory (ROFT) [9, 11, 13]. It is now well established experimentally and by simulation the existence of the dynamic correlation length scale that are associated to the dynamical heterogeneity present in glassy materials [9, 11]. However, scientists are now more concentrated on using DH to understand and predict glass transitions. The are a limited number of experiments confirming the presence of DH, and this is due to the difficulty to probe at such short spatial and temporal resolution [12, 14, 15]. On the other hand, molecular simulation techniques are the ideal method to obtain the microscopical description of the sample dynamics around the vitrification process[2, 4, 6, 9–11, 16]. Many tools have been combined with statistical results by molecular simulations to infer information on the slow dynamics process with the aim of understanding and predicting glass-forming properties from the sole input of microscopical structural description [17–21]. Among those tools, many recent molecular simulation studies involves Machine Learning [18, 19, 21] and neural networks techniques [17, 20].

1.1 Motivation and objectives of the thesis

This section present a list of the topics that have been investigated with the complementary objectives of this thesis.

- Different states of matter of a colloidal model have been generated and characterized using MD simulations [22]. In this preliminary model-study, Glass-properties have been reproduced in excellent agreement with previ-

ous simulation studies [16].
- Structural information have been collected in order to characterize the dynamic relaxation of the model in the liquid and glassy state.
- Dynamical heterogeneity associated with a vitreous state have been studied using standard methods, and, the description have been enriched using a basic machine learning approach.

1.2 Thesis Description

In this thesis, the focus is directed towards the relaxation properties and dynamic heterogeneity of a glass forming material using Molecular Dynamics (MD) simulations [23]. Furthermore, a machine learning techniques, i.e. k-means clustering [24], is used to infer physical interpretation from many statistical data. This thesis is organized as following. Chapter 2 presents briefly the theory necessary to understand the vitrification problem, Mode Coupling Theory, and, the previous works related of the topics of this study. Chapter 3 focuses on the methods that have been used to study the glass formation and the relaxation process. The chapter explains the basic of the molecular simulations technique, the approach to study the particles' mobility and the machine learning tool used to complete the study on the dynamic heterogeneity. In Chapter 4, all the results are presented, and, finally in Chapter 5 conclusions and further works are discussed.

Chapter 2

Theory

In this chapter, several important concepts related to the investigation of the relaxation process and the dynamic heterogeneity analysis is presented. The principles and equations in this chapter are vital for understanding many of the results which will be discussed in chapter 4.

2.1 Phenomenology of glass formation

Despite the enormous implication, the understanding of the glass transition is still lacking. In this subsection structural, dynamical and thermodynamics properties will be considered in the context of vitrification, and, for our purposes, it is helpful to enunciate some of the salient features of a glass-forming system [4]:

- Presence of disorder and/or frustration in the structure, and, very weak dependence of structural properties on external parameters, like temperature and pressure.
- Strong dependence of dynamical properties on external parameters.
- Non exponential time-decay of the relaxation at low temperature.
- Presence of 'Dynamic Heterogeneity' (DH), or spatially heterogeneous nature of the local dynamic.

From a phenomenological point of view, the systems undergoing a glass transition, assume a disordered or "amorphous" structure in the long range, and, a short range order that only involve local regions presenting very different heterogeneous features between them. Furthermore, a rapid slowing down of the dynamics, the sudden rise of viscosity and the appearance of rigidity are all appearing during the glass transition. The dynamics of glasses is known to be very sluggish and heterogeneous, and it occurs in a wide variety of systems, i.g. Polymeric liquids, granular material, such as powder and sand, gels, and foams and also metallic alloys. Dynamical heterogeneity has also been found in biological systems, e.g. in collective cell migration, cancer invasion, and wound healing. Structural and dynamical properties, as well, as evidence of dynamical heterogeneity can be measured by experiments as well as simulations [4, 9, 11]. For example, the temperat-

ure dependence of dynamical quantities (e.g. viscosity, η, diffusion constant of the atoms, D, and relaxation time, τ), or structural (e.g. arrangements and bonding of atoms, patterns and boundaries) and thermodynamics properties (volume, enthalpy, heat capacity etc...) are routinely measured. The main conclusions of many experiments and simulations [4, 6], show that most liquids can be 'supercooled' in the meta-stable glassy-regime characterized by structural and thermodynamic properties, that depend **weakly** on the temperature, and, by dynamic quantities, like η, τ and D, that depend **strongly** on the temperature. As a consequence, vitrifying a material upon cooling (or compression) results into a sudden dramatic increase of the viscosity but only minute structural changes.

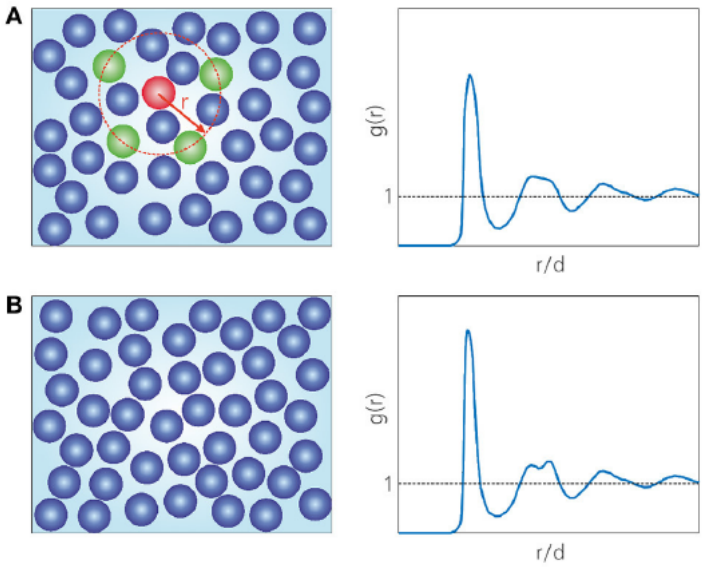


Figure 2.1: Schematic pictures of the structure of a liquid (A) and a glass (B). On the left panels, particles' arrangement is shown for the two state, while on the right panels, the corresponding radial distribution function g(r) are shown. Four green particles distant a certain radius r from a reference red particle are shown in the top left panel. Pictures from [6].

On left panels of Figure 2.1, it can bee seen that the particles' arrangements of the liquid, (A), and the glass, (B), states are very similar. It is also instructive to calculate the radial distribution function $g(r)$ that represents the average distribution of atoms around any given atom within the system. The right (B) panel clearly shows the typical short range order and absence of long range order that is characteristic of the corresponding structure in the liquid state (right (A) panel).

In Figure 2.2 a scheme of a typical 'Angell-plot' is shown. The plot shows the temperature dependence of the viscosity for different glass-forming materials. T_g represents the vitreous transition temperature. The sudden increase of viscosity is evident for all the materials, and, it is also clear that there is not a universal behavior close to the vitreous transition suggesting more than one underlining

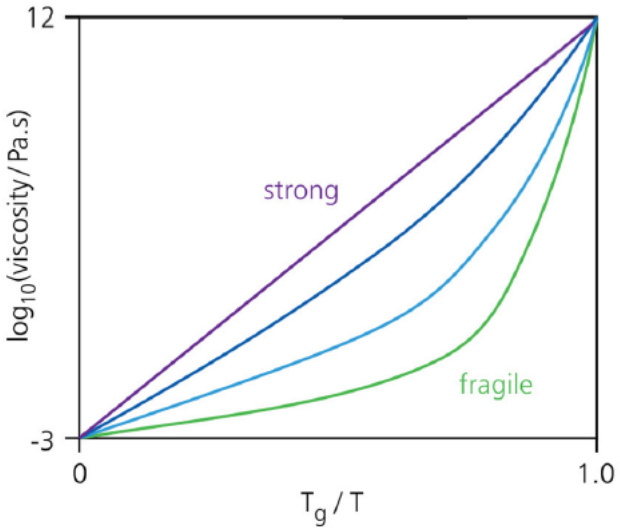


Figure 2.2: Schematic picture of a typical 'Angell- plot' taken from reference [6]. The logarithm of the viscosity is shown in function of the inverse of the Temperature, T, and normalized respect to the glass transition Temperature, T_g. The *strong* and *fragile* materials are represented in the graph.

mechanism of vitrification. All the curves at high temperature in the diagram, can be fitted with an Arrhenius-kind law:

$$\eta(T) = \eta_0 \exp(E(t)/(k_\beta T)). \tag{2.1}$$

However, for certain systems, i.e. *Strong* glass-forming materials, the activation energy $E(t)$ is constant for all range of temperature, while for other systems, i.e. *fragile* glass-forming materials, only the local slope of the curve can be interpreted as activation energy $E(t)$, because this appears to change and increase rapidly with the decreasing of the temperature T. In the *fragile* curves, in the region of intermediate values for T_g/T, it appears a bending that becomes more and more pronounced at $T_g/T \approx 0.7$. The differences between those two broad classes of materials is not very clear at microscopic level. However, it is a fact that strong glass-formers are characterized by open network, i.e. SiO_2, whereas the structure of fragile materials is more compact. Furthermore, there are many empirical correlation between fragile materials and structural properties [4].

The viscosity response to the decrease of temperature is clearly an effect of the slowing down of the dynamics, and many scientists have tried to understand if the other typical time-scales of the systems, such as diffusion or relaxation times, have the same temperature dependence. The super-Arrhenius Vogel-Fulcher-Tammann law (VFT) 2.2, [25–27] is able to fit some experimental viscosity-temperature data, but not all, at intermediate η. Although this formula is a good representation of certain data, it has not theoretical foundation, furthermore, there is experimental evidence that the VFT law does not hold exactly [28].

$$\eta(T) = \eta_0 \exp(A/(T - T_0)). \tag{2.2}$$

However, a nice feature of the VFT law is that all the mentioned timescales can be represented by a similar formula:

$$X(T) \propto \exp(C_X T_0/(T - T_0)). \tag{2.3}$$

In Equation 2.3, X represents η, D^{-1}, or τ, while C_X, T_0 and A are empirical material-dependent fitting parameters. It should be noted that Equation 2.3 can be reduced easily to Equation 2.2 for the viscosity case. T_0 is the so called "Vogel-Temperature", and it is usually around 50° lower than the vitrification temperature.

On the other hand, the Mode Coupling Theory [29–31] has shown great success in predicting the experimental data of certain materials at intermediate values of η, i.e. $10^{-1}P \leqslant \eta \leqslant 10^2 P$.

$$\eta(T) = \eta_0 (T - T_c)^{-\gamma}. \tag{2.4}$$

The MCT is arguably the most successful theory of glass-transition and it is based entirely on first principles [6, 30]. MCT aims to predict the microscopic relaxation of glass-forming materials using the structure of the material as input, i.e. static structure function. Both, scattering experiments and MD simulations [2, 4, 6, 8] provide a direct measure of the so called static scattering function, $S(\vec{k})$, which is related to the intermediate scattering function, $F(\vec{k}, t)$, through fourier transform. Before giving a brief introduction of the MCT, the scattering functions and their importance will be explained .

2.2 Static structure factor

An important tool used in soft matter physics to describe the structure of a system, is the static structure factor. For a system with N particles confined in a volume V, the density profile is given as $\rho(\vec{r}) = \sum_{i=1}^{N} \delta(\vec{r} - \vec{r}_i)$. In the reciprocal space, the density function is given as

$$\hat{\rho}_{\vec{k}} = \int \rho(\vec{r}) e^{i\vec{k}\cdot\vec{r}} d\vec{r} = \sum_{i=1}^{N} e^{i\vec{k}\cdot\vec{r}_i}. \tag{2.5}$$

Note that the wavevector $\vec{k}$ is defined through the physics convention, of which the magnitude is the wave-number $k = \mid \vec{k} \mid = \frac{2\pi}{\lambda}$. The static structure factor is now defined as the Fourier transform of the density-density correlation function[7]

$$S(\vec{k}) \equiv \frac{1}{N} \left\langle \hat{\rho}_{\vec{k}} \hat{\rho}_{-\vec{k}} \right\rangle. \tag{2.6}$$

By inserting (2.5) into (2.6), we get

$$
\begin{aligned}
S(\vec{k}) &= \frac{1}{N}\left\langle \sum_{i=1}^{N} e^{i\vec{k}\cdot\vec{r}_i} \sum_{j=1}^{N} e^{-i\vec{k}\cdot\vec{r}_j} \right\rangle \\
&= \frac{1}{N}\left\langle \sum_{i=1}^{N}\sum_{j=1}^{N} e^{i\vec{k}\cdot(\vec{r}_i-\vec{r}_j)} \right\rangle \\
&= \frac{1}{N}\left\langle \sum_{i=1}^{N}\sum_{j=1}^{N} \cos\left(\vec{k}\cdot(\vec{r}_i-\vec{r}_j)\right) + i\sum_{i=1}^{N}\sum_{j=1}^{N} \sin\left(\vec{k}\cdot(\vec{r}_i-\vec{r}_j)\right) \right\rangle, \qquad (2.7)
\end{aligned}
$$

which is very useful in computer simulations because it requires only the position of the particles in the system to be calculated. In most cases, the static structure factor as a function of magnitude of the wave vector $|\vec{k}|$ is of more importance. $S(k)$ is then calculated by averaging over a set of k-vectors with the same magnitude and equally spaced in a spherical shell in reciprocal space.

Furthermore, if the system contains two types of particles, the static structure factor will depend on which particle correlations is being considered. Thus, for a binary system the partial static structure is given as

$$
S_{\alpha\beta}(\vec{k}) = \frac{1}{N}\left\langle \sum_{i=1}^{N_\alpha}\sum_{j=1}^{N_\beta} \cos\left(\vec{k}\cdot(\vec{r}_i-\vec{r}_j)\right) + i\sum_{i=1}^{N_\alpha}\sum_{j=1}^{N_\beta} \sin\left(\vec{k}\cdot(\vec{r}_i-\vec{r}_j)\right) \right\rangle, \qquad (2.8)
$$

with $\alpha, \beta = A, B$ as the two possible particle types.

A useful property of the static structure factor is its relation to the scattering patterns obtained in various diffraction experiments. Because the static structure factor describes how a system scatter incident radiation, one can do an X-ray diffraction experiment on a real system to validate the configurations obtained through molecular dynamics simulations.

2.3 Intermediate scattering function

The intermediate scattering function is a time-dependent correlation function that can be obtained by statistical calculations using molecular dynamics simulation [2, 4, 6], and, it reveals important information on the microscopic nature of the vitrification and relaxation process.

In order to introduce the intermediate scattering function, the Van Hove function first needs to be defined. The Van Hove correlation function for a system of N particles is given as

$$
G(\vec{r}, t) = \frac{1}{N}\left\langle \sum_{i=1}^{N}\sum_{j=1}^{N} \delta\left(\vec{r} + \vec{r}_j(0) - \vec{r}_i(t)\right) \right\rangle, \qquad (2.9)
$$

where $\vec{r}_i(t)$ is the time dependent position of particle i. It is often useful to separate the sums into two parts by distinguishing between the sums where $i = j$ and $i \neq j$

$$G(\vec{r},t) = \frac{1}{N}\left\langle \sum_{i=1}^{N} \delta\big(\vec{r} + \vec{r}_i(0) - \vec{r}_i(t)\big)\right\rangle + \frac{1}{N}\left\langle \sum_{i=1}^{N}\sum_{j\neq i}^{N} \delta\big(\vec{r} + \vec{r}_j(0) - \vec{r}_i(t)\big)\right\rangle$$
$$\equiv G_s(\vec{r},t) + G_d(\vec{r},t). \tag{2.10}$$

Here $G_s(\vec{r},t)$ is conventionally known as the self-part of the Van Hove function and describes the average motion of a particle starting at a given position. In other words, $G_s(\vec{r},t)$ gives the probability for finding a particle at a position $\vec{r}$ away from its starting position after a time t. $G_d(\vec{r},t)$ is known as the distinct-part of the correlation function and describe the average motion of the remaining $(N-1)$ particles. Because the distinct-part is calculated relative to the initial position of particle i, $G_d(\vec{r},t)$ corresponds to the possibility of finding one of the other particles at a distance $\vec{r}$ in relation to particle i. In the limit $t \to \infty$, both correlation functions become independent of $\vec{r}$ as $G_s(\vec{r}, t\to\infty) \sim 1/V$ and $G_d(\vec{r}, t\to\infty) \sim \rho$.

The Van Hove function considers particle correlations in space. Similarly, the correlations can be studied in the reciprocal space by analyzing the Fourier components of the Van Hove function. The intermediate scattering function is defined as the Fourier transform of the Van Hove function [7]

$$F(\vec{k},t) = \int G(\vec{r},t) e^{-i\vec{k}\vec{r}t} d\vec{k}. \tag{2.11}$$

Similarly to the Van Hove function, $F(\vec{k},t)$ can be separated into a self-part $F_s(\vec{k},t)$ and a distinct-part $F_d(\vec{k},t)$,

$$F_s(\vec{k},t) = \int G_s(\vec{r},t) e^{-i\vec{k}\vec{r}t} d\vec{k}, \tag{2.12}$$

$$F_d(\vec{k},t) = \int G_d(\vec{r},t) e^{-i\vec{k}\vec{r}t} d\vec{k}. \tag{2.13}$$

By inserting the Van Hove functions from (2.10) into (2.12) and (2.13), the intermediate scattering functions can be calculated directly based on the particle trajectories,

$$F_s(\vec{k},t) = \frac{1}{N}\left\langle \sum_{i=1}^{N} \exp[i\vec{k}(\vec{r}_i(t) - \vec{r}_i(0))]\right\rangle, \tag{2.14}$$

$$F_d(\vec{k},t) = \frac{1}{N}\left\langle \sum_{i=1}^{N}\sum_{j=1}^{N} \exp[i\vec{k}(\vec{r}_j(t) - \vec{r}_i(0))]\right\rangle. \tag{2.15}$$

Similarly to the static structure factor, it is often more convenient to calculate $F_s(\vec{k},t)$ for a specific magnitude of $\vec{k}$,

$$F_s(k,t) = \frac{1}{N}\left\langle \sum_{i=1}^{N} \exp[ik|\vec{r}_i(t) - \vec{r}_i(0)|]\right\rangle. \tag{2.16}$$

The intermediate scattering function considers the correlation between particles' position in an initial configuration and the same particles' position after a time t. Thus, $F(\vec{k}, t)$ can be used to investigate the relaxation time of the system. A sketch of a typical behavior for $F_s(\vec{k}, t)$ in a liquid state, high temperature, and a glassy state, low temperature, is shown in figure 2.3. For both, liquid and glassy state, at short timescales, the particles have not been given enough time to interact with each other, and, as a consequence their displacements is very limited. This initial time period is often referred as ballistic regime and, the time dependence of $F_s(\vec{k}, t)$ is found to be quadratic in time [32].

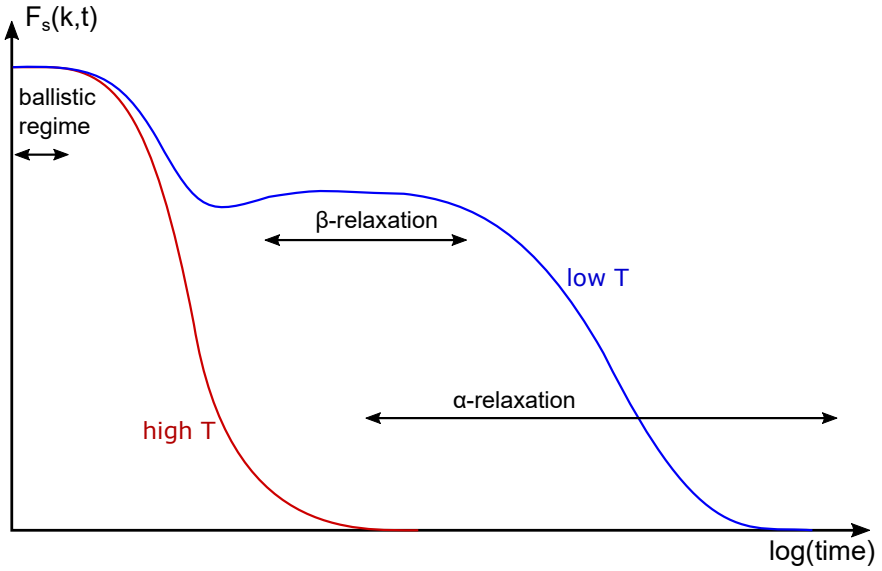


Figure 2.3: Sketch of the self-part of intermediate scattering function for a liquid state (red) and a glassy state (blue).

After this initial regime, at high temperatures, the particles in the system are expected to have a diffusive motion and collide rather freely. Thus, the correlation with respect to the initial configuration breaks down fast, meaning that the system lose the "memory" of the previous configuration. This phenomena appears in the time-dependence of $F_s(\vec{k}, t)$ as a fast time-exponential decay described as Debye relaxation, also called short relaxation regime.

The Kohlrausch-Williams-Watts function (KWW) below is well able to fit the $F_s(\vec{k}, t)$ [2, 4].

$$F_s(t) = A\exp\left(-(t/\tau)^{\beta}\right), \tag{2.17}$$

In equation 2.17, A is a fitting parameter called amplitude, τ is the parameter that can describe the relaxation time, and β is defined as the KWW-exponent. $\beta > 1$ in a typical Debye-relaxation indicates a diffusive, Gaussian motion of the particles.

For the glassy state, lower temperature in the sketch, the evolution of the $F_s(\vec{k}, t)$ is different. After a ballistic regime, at intermediate times, the particles

start to influence each other and the dynamic of the system is said to be sluggish or slow. This is represented in $F_s(\vec{k}, t)$ as the plateau regime, where some of the "memory" from the initial configuration is conserved, resulting in what is known as cage-effect [2, 4]. Many of the particles are nearly frozen and the surrounding particles can effectively act like a cage. The cage-effect is illustrated in figure 2.4. With most of the particles trapped and unable to diffuse, the correlation to the initial configuration is partially conserved, and, a particular slow relaxation process, i.e. β-relaxation, occurs. Eventually, a complete relaxation is reached at much longer times. In this regime the particles "escape" the cage and eventually a fast relaxation regime take place. The α-relaxation regime is related to the complete breakdown of the correlation. This regime is characterized by a non-Debye time-exponential decay of the $F_s(\vec{k}, t)$ also called stretched-exponential regime. Equation 2.17 can be used to fit the α-relaxation regime, and, it is characterized by a longer relaxation times ($\tau \approx \mu s$) and a β exponent smaller respect to the liquid counterpart. In particular, a typical $\beta < 1$ exponent in the glassy state indicates a non-Gaussian motion.

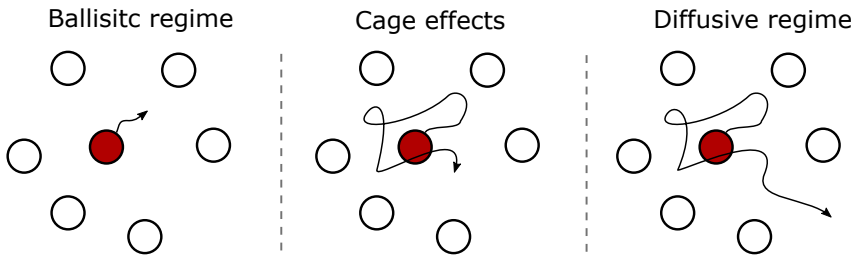


Figure 2.4: Illustration of particle movements for different regimes occurring during simulation of a glassy state.

The origins and cause of the stretched-exponential decay are not well known yet [4]. Among many hypothesis, the cause of this behavior is connected to the spatial heterogeneity that appears during the β- relaxation of the glassy system. This heterogeneity or spatial local variation, is such that each particle experiences a different decay process, and in turn, in different relaxation times, that all combined, give raise to the stretched-exponential time-dependence [2]. The Mode Coupling Theory is able to predict the features of the intermediate scattering function and the glass transition taking as input the sole structure function.

2.4 Mode coupling theory

As already stated, the Mode Coupling Theory is based on first principle, i.e. it does not rely on any phenomenological assumption, and, successfully predicts the full microscopic dynamics of the glass transition starting solely from the time-independent structural information defined by the static structure function.

The mode-coupling theory (MCT) was developed in the seventies and evolved

along the years [29, 30], and includes complex equations to describe the dynamics of simple fluids. The equations proved a semi-quantitative description of the dynamics at the glass transition, and, importantly it predicts a glass transition sometime in good agreement with experiments and simulation results [33, 34].

In particular, MCT provides a equation for the intermediate scattering function that can be solved numerically giving the static structure function as input. From that information, MCT manages to successfully predict all the microscopic relaxation dynamics for the glass-forming material as a function of wave-vector, temperature, and, density.

The most prominent successes are the qualitative prediction of the glass transition, the cage effect, and, in general, all the scaling behaviour of the intermediate scattering function. The MCT is not quantitatively accurate, and does not include the concept of fragility, nor the Dynamical heterogeneity, nor the violation of the Stokes-Einstein equation. There are different extension of the theory (Generalized MCT and Inomogeneous MCT) that indeed offer a more quantitative prediction of the intermediate structure function and a general framework for the DH using some approximation [6, 9].

MCT is not directly used in this thesis, but the theory provides the glass transition temperature for the simulation model which is a reference in all my discussions.

2.5 Dynamic heterogeneity

The occurrence of spatially heterogeneity in glassy dynamics, commonly referred to as dynamic heterogeneity, is well established in both experimental and numerical studies [35, 36]. The appearance of those local heterogeneous regions happens in the cage-regime, when some particles are undergoing a rattling motion inside the 'cage' formed by the surrounding particles. Those local regions are correlated into clusters that have their own non-exponential decay resulting into a non Gaussian motion or distribution of displacements the we have described in the previous section [2]. The common way of describing this heterogeneity is to consider the displacements of the particles in time, i.e., the Van Hove function, and compare it to a Gaussian distribution. In the glassy state and at intermediate times, the distribution of displacements shows a large tail indicating the presence of highly mobile particles. At his stage, the displacement distribution deviate from the Gaussian, diffusive form. Technical aspect of the method to reveal DH will be described in section 3.3.1.

The parameter that measure the presence of dynamic heterogeneity, i.e. the deviation of the displacements' distribution from the Gaussian distribution is the non-Gaussian parameter α_2 [9], which is defined as

$$\alpha_2(t) = \frac{2\langle|\vec{r}(t)|^4\rangle}{5\langle|\vec{r}(t)|^2\rangle^2} - 1. \tag{2.18}$$

Here $|\vec{r}(t)|$ is the magnitude of the displacement of a particle in the system. The

α_2 parameter indicates which temperatures and times give presence of DH. In particular, the α_2 parameter peak measure the time at which the degree of heterogeneity reaches a maximum [9].

Another popular instrument used to measure the dynamic heterogeneity in the systems, is the four-point susceptibility χ_4 that is calculated from the fluctuations of the intermediate scattering function $F_s(k,t)$. Spatial heterogeneity in the samples give rise to inconsistencies in the function $F_s(k,t)$, and, this function reveal the variance of these fluctuations [37, 38],

$$\chi_4(t) = N\Big[\langle F_s(k,t)^2\rangle - \langle F_s(k,t)\rangle^2\Big]. \tag{2.19}$$

N is the total number of particles and the wave number k is chosen by the peak of the static structure factor. Further, $\langle ...\rangle$ determines a thermal average.

The area and the peak of this function is also a measure of the size of the DH that are cooperatively rearranging in time. The increase of this peak value corresponds to a growing correlation length or clusters size of this heterogeneous local regions. In addition, the position of its maximum is proportional to the α-relaxation time τ found for the same system. Analyzing how the susceptibility χ_4 evolves over time can accordingly indicate how the heterogeneity grows as the temperature is lowered. Typical behavior of χ_4 in colloidal systems has shown an increase of the peak value as the temperature is lowered down below the transition temperature [9].

Chapter 3

Methods

The content of the following sections is devoted to introducing the methods used to analyze the binary colloidal Yukawa system of this thesis. The first part is a general introduction to the field of molecular dynamics simulations in order to familiarize the reader to the different concepts and ideas regarding this method for simulating materials. Additionally, the exact system and dimensions used for the system is described in more detail here. The second part of this chapter describes the method used to study the dynamical heterogeneities and includes an introduction to k-means and the capabilities of this tool.

3.1 Molecular Dynamics simulations

Molecular dynamics (MD) is a numerical simulation method used to analyze the structure and trajectories of particles in systems with sizes ranging from only a few particles to billions of particles. As a scientific tool, MD has had huge success in the last decades in fields like soft matter physics and fluid dynamics due to its simplicity, low cost and an exponential increase of hardware speed-ups offered by Moore's Law and large-scale parallelism. The dramatic power increase for computations has allowed for bigger and more accurate simulations which has increased the relevance of results gathered through MD. Additionally, a pure digital experimentation method allows for analysis on specific or abstract systems which can be impossible to replicate in physical experiments. With all the trajectories in the system directly accessible, measurements like energy, pressure, temperature, etc. can easily be calculated and controlled without being dependent on expensive and complex experimental equipment. Although the advantages are many, numerical MD inherently rely on approximations. Thus, any results originated from this field have some limitiations and should be viewed as preliminary experiments performed with computers.

The main task in molecular dynamics is to numerically integrate a set of coupled differential equations based on Newton's equations of motion [39]. For a system containing N particles, these are given by

$$\begin{aligned} \vec{F}_i &= m_i \cdot \frac{d\vec{u}_i}{dt} \\ \vec{u}_i &= \frac{d\vec{r}_i}{dt}, \end{aligned} \qquad i = (1, 2, ..., N) \tag{3.1}$$

where $\vec{r}_i$ and $\vec{u}_i$ are the position and velocity of a particle i respectively. m_i is its mass and $\vec{F}_i$ is the net force acting on the particle. The forces acting on a particle can be pairwise interactions, many-body interactions or external forces acting on the system, e.g. electromagnetic forces.

3.1.1 The numerical integration method

To perform a molecular dynamics simulation, an initial configuration is needed. This initial configuration of particles could be critical to the behavior of the system that the user would want to investigate. For instance, the initial configuration defines the packing density of particles and the mobility at low temperatures. Further it can be important to ensure a sufficient initial distance between the particles, especially for repulsive interactions. It is thereby common to arrange the system on a lattice instead of choosing the positions at random. Next, all particles are given a velocity which is shifted such that the total momentum over the system is zero. Sometimes, in order to enforce a specific temperature T in the system, the velocities are scaled such that the mean kinetic energy of the system matches the desired temperature T through the equipartition theorem. Another simpler method is however to initialize the system at zero temperature and gradually increase the temperature during the equilibration simulation, which then will have a slightly longer duration.

A crucial part of molecular dynamics is to determine the forces acting on all the particles in the system. In some cases, only pairwise additive interactions between the particles are modeled. Thus, it is possible to describe the forces through a pair-potential, $\phi(\vec{r})$. With a system of N particles, this computation over the forces of the system is $\mathcal{O}(N^2)$, which makes it one of the most computational heavy parts in MD. There are many common ways to reduce this computation. In general, for a decaying pair-potential, the contributions from particles that are relatively far away can be neglected compared to the interactions from nearby particles. By first evaluating all the pairwise distances in the system, one can introduce a cut-off at r_{cut}, where all interactions operating further apart than this distance, is neglected [40]. The cut-off distance r_{cut} should be chosen with respect to the given pair-potential and particle density of the system. Conveniently, this approximation allows for a further speed increase in the calculations as one can introduce a neighbor list to each particle. (cite Lammps article) These lists will contain information on which pairs to evaluate on each step to avoid calculating distances between particles that is expected to be too far apart, $|\vec{r}| \gg r_{cut}$. Unless the temperature is high, these neighbor lists don't need to be updated too often and a significant reduction in computational cost is attained for large systems.

With the pairwise forces in the system determined, the next part is to integrate the Newton's equation of motion (3.1). There are several ways to do this and in this study one of the most common ways was used, namely the velocity-Verlet scheme. The velocity-Verlet is based on a Taylor expansion of the position of a given particle i at a time $t \pm \Delta t$,

$$\vec{r}_i(t \pm \Delta t) = \vec{r}_i(t) \pm \frac{\partial \vec{r}_i(t)}{\partial t}\Delta t + \frac{1}{2}\frac{\partial^2 \vec{r}_i(t)}{\partial t^2}\Delta t^2 \pm \frac{1}{6}\frac{\partial^3 \vec{r}_i(t)}{\partial t^3}\Delta t^3 + \mathcal{O}(\Delta t^4). \quad (3.2)$$

By adding the forward and backward form of equation (3.2) and substituting in the force $\vec{F}_i(t)$ for the acceleration, the basic Verlet scheme is given as

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\vec{F}_i(t)}{m_i}\Delta t^2 + \mathcal{O}(\Delta t^4). \quad (3.3)$$

Similarly, by subtracting the two forms of the equation, some rearranging gives us the velocity $\vec{u}_i(t)$,

$$\vec{u}_i(t) = \frac{\vec{r}_i(t + \Delta t) - \vec{r}_i(t - \Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2). \quad (3.4)$$

Equations (3.3) and (3.4) form what is known as the basic Verlet algorithm. Because the equations include $\vec{r}_i(t - \Delta t)$, the algorithm is not self-starting. A solution is to solve (3.4) with respect to $\vec{r}_i(t - \Delta t)$ and then inserting it into (3.3), which yields

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{u}_i(t)\Delta t + \frac{\vec{F}_i(t)}{2m_i}\Delta t^2 + \mathcal{O}(\Delta t^3). \quad (3.5)$$

An expression for the velocity is given by using (3.4) to define $\vec{u}_i(t + \Delta t)$ and then substituting $\vec{r}_i(t + 2\Delta t)$ by using (3.3). Some rearranging then gives a new expression for the velocity of a particle i,

$$\vec{u}_i(t + \Delta t) = \frac{\vec{r}_i(t + \Delta t) - \vec{r}_i(t)}{\Delta t} + \frac{\vec{F}_i(t + \Delta t)}{2m_i}\Delta t + \mathcal{O}(\Delta t^2). \quad (3.6)$$

Inserting (3.5) into (3.6) finally gives a simple expression for updating the velocity,

$$\vec{u}_i(t + \Delta t) = \vec{u}_i(t) + \frac{1}{2m_i}\big(\vec{F}_i(t + \Delta t) + \vec{F}_i(t)\big)\Delta t + \mathcal{O}(\Delta t^2). \quad (3.7)$$

Equation (3.5) and (3.7) represent the algorithm known as the velocity-Verlet and is used on all the particles in the system to update their position and velocity for each iteration. From the equations one can see that the global error associated with the velocity-Verlet scheme is of the third order in the position and second-order in the velocity.

A canonical ensemble was used in this study to ensure a system environment with a controlled temperature. This means that the number of particles, N, the

volume, V, and temperature, T, of the system was conserved during a whole simulation run. A simulation like this is usually referred to as an NVT simulation because of the constraints set on three of its parameters. A micro canonical ensemble, NVE, could also have been chosen for this study, but NVT-simulations controls the sample's temperature making it comparable with a real experiment situation. To ensure the constant temperature, the system was coupled to a thermostat. The thermostat exchange energy with the system to, among other things, prevent viscous heat to change the environment of the simulations. The exact thermostat used in this thesis is described in the next subsection.

A software called LAMMPS [41] was used to do the MD simulations in this project. LAMMPS stands for Large-scale Atomic/Molecular Massively Parallel Simulator and is designed to run efficiently on parallel computers. It's possibility to customize the simulation system with specific atom types, force fields and boundary conditions, made it easy to incorporate the colloidal system to be studied in this project. An example of the input files used in the software is shown in Appendix A for the initial configuration of the system simulating at a temperature of $T = 0.14$.

Although LAMMPS make molecular dynamics simulation efficient, the computational task is still cumbersome and require good hardware to simulate systems of adequate size and time. Fortunately, access to the Swiss National Supercomputing Centre (CSCS) was granted and all MD simulations performed for this thesis was run on their HPC cluster named Piz Daint. Having one of the most powerful supercomputers in the world [42] and its GPU accelerated nodes, allowed for gathering of MD data that capture the inherently slow dynamics of the glassy systems studied in this work.

3.1.2 The system

The specific system used in this study was a binary colloidal system containing a 50:50 mixture with a total of $2N_A = 2N_B = N_{tot} = 16000$ particles of kind A and B. The diameters of the colloids are given as $d \equiv d_{AA} = 1.0$, $d_{AB} = 1.1d$ and $d_{BB} = 1.2d$, and their masses are given as $m = m_A = m_B = 1.0$. Hence the only difference between the two types of particles, is their diameter. The particles are arranged into a cubic sample of density $\rho = 0.675 m_A/d^3$. Note that all quantities are defined to be unitless, which often is referred to as reduced units. In practice, this is done by setting the fundamental quantities mass, σ, ϵ and the Boltzmann constant k_B equal to one. With this as a base, all other observables are constructed using these fundamental quantities.

The reason for the choice of having a binary colloidal system instead of a single type of colloids, is to prevent crystallization of the sample when quenched. Preparing a glass in molecular dynamics simulation can be difficult, and glass-forming liquids need to be engineered carefully. For systems with a single type of particles, the configuration crystalizes when quenched below the melting point. This is because the internal energy minima of the particles coincide at the same

point. This is however not true for the binary case, and thus the crystal structure is not as strong and harder to construct.

For the force field, a potential called the Yukawa potential was used to model the interactions between the particles. The Yukawa potential is named after the physicist Hideki Yukawa and is a pairwise potential used to simulate the interactions in systems consisting of attracting or repelling particles. The potential is given by

$$V_{\alpha\beta}(r) = \epsilon_{\alpha\beta} d_{\alpha\beta} \frac{\exp(-\kappa_{\alpha\beta}(r - d_{\alpha\beta}))}{r}, \tag{3.8}$$

with $\alpha, \beta = A, B$. The energy parameters $\epsilon_{\alpha\beta}$ was set to $\epsilon \equiv \epsilon_{AA} = 1.0$, $\epsilon_{AB} = 1.4\epsilon$ and $\epsilon_{BB} = 2.0\epsilon$, and the screening parameters $\kappa_{\alpha\beta}$ was set equal with a value of $\kappa_{AA} = \kappa_{AB} = \kappa_{BB} = 6/d$. A plot of the potential with the specific values are shown in figure 3.1. From this plot and equation (3.8), one can see that the potential is strictly positive and that the particles contained in the system are always repelling each other.

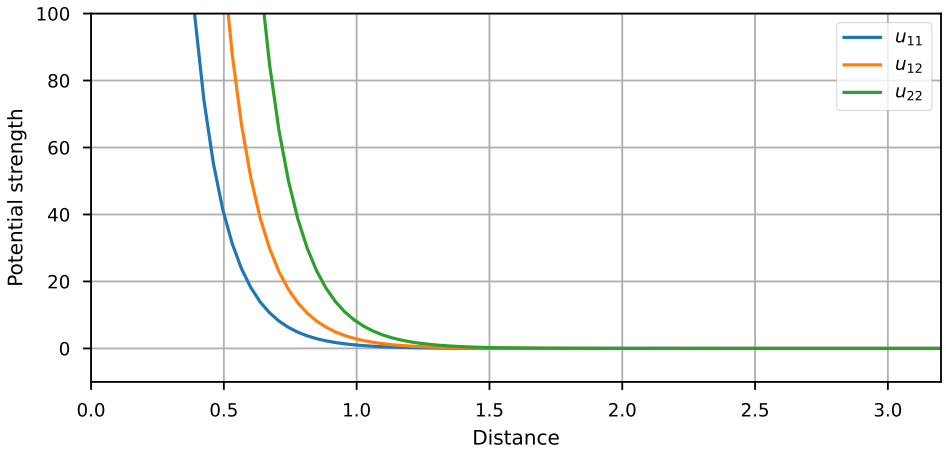


Figure 3.1: A plot of the specific Yukawa potential used in the simulations.

The potential is often referred to as the screened Coulomb potential as it can be reduced to the Coulomb potential simply by letting the screening parameters be zero. A predefined module for the Yukawa potential was used in the Lammps simulation software.

In this study, the so called Dissipative Particle Dynamics (DPD) thermostat was coupled to the system [43]. The DPD thermostat can be formulated by the equations of motion given as

$$\frac{d\vec{r}_i}{dt} = \frac{\vec{p}_i}{m_i}, \qquad \frac{d\vec{p}_i}{dt} = \vec{F}_i + \vec{F}_i^D + \vec{F}_i^R, \tag{3.9}$$

where $\vec{r}_i$ is the position and $\vec{p}_i$ is the momentum of particle i. $\vec{F}_i$ is the conservative net force on particle i given by the gradient of the Yukawa potential formulated

in (3.8) for all particles pairs including this particle. The thermostat works by adding a dissipative force, $\vec{F}_i^D = \sum_{j \neq i} \vec{F}_{ij}^D$, and a random force, $\vec{F}_i^R = \sum_{j \neq i} \vec{F}_{ij}^R$, to the equation of motion.

The dissipative force is defined as

$$\vec{F}_{ij}^D = -\zeta \omega^d(r_{ij})(\hat{r}_{ij}\vec{u}_{ij})\hat{r}_{ij}, \tag{3.10}$$

where $\vec{u}_{ij} = \vec{u}_i - \vec{u}_j$ is the relative velocity between particle i and j. $\hat{r}_{ij}$ is the unit vector in the interatomic axis $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$. ω^D denotes a weight function which vanishes for $r \geqslant r_c = 1.25d$. The friction constant ζ was set to a value of $\zeta = 12$ to give the system microscopic properties close to that of pure Newtonian dynamics, $\zeta = 0$. This means that the term $\vec{F}^D$ describes the frictional forces in the system due to interactions between the particles and the use of relative velocities ensures that the momentum is locally conserved. The random force is given in a similar way by

$$\vec{F}_{ij}^R = \sigma \omega^R(r_{ij})\theta_{ij}\hat{r}_{ij}, \tag{3.11}$$

with σ as the noise strength, which was set equal to $\sigma = \sqrt{2k_b T\zeta}$ by using the fluctuation-dissipation theorem. θ_{ij} is a Gaussian white noise variable and ω^R is another weight function simply given as $[\omega^R(r)]^2 = \omega^D(r)$. The described thermostat was implemented using an internal LAMMPS command, which can be seen in appendix A.

All parameters in this system are carefully chosen to allow for glass-formation when the system is quenched down below the transition temperature, i.e., no crystallization or phase separation of the sample. This has been showed for the same model in previous studies [16], as well as the pre-study work of this thesis [22]. In the pre-study work, the exact same system was investigated by characterizing the different states of matter through analyzing the structure and diffusivity of the samples. Furthermore, the non-Newtonian behaviors of the glassy state was reproduced with excellent agreement with previous studies.

3.2 Initialization of the system

To initialize the system, the particles was placed into a body-centered cubic (bcc) lattice configuration. The unit cell of the bcc structure hold two particles, this makes it an inherent binary system if the two particles are differentiated. In this case one was given a diameter of d_{AA} and the other was given a slightly bigger diameter of d_{BB}. The lattice parameter was set to give the desired density and the resulting colloidal system is visualized in figure 3.2.

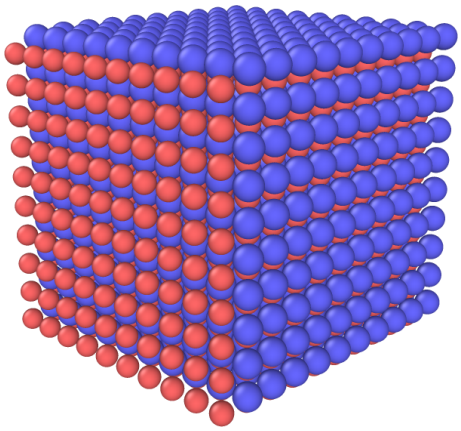


Figure 3.2: Visualization of the initial body-centered cubic configuration of the binary colloidal system.

The Yukawa potential used in this study easily forms a disordered or glass-like state if the initial configuration of the particles is not well ordered, even without the usually required quenching process. The bcc structure was therefore a good choice for the initial configuration for this study because it give a high level of order in the initial structure, which can be confirmed by the distinct peaks of the radial distribution function given in figure 3.3.

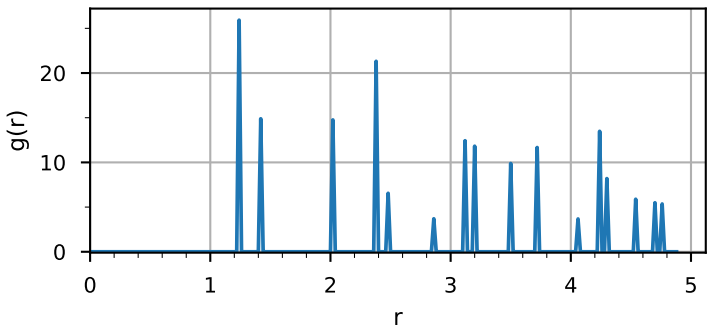


Figure 3.3: Radial distribution function of initial bcc structure.

The initial configuration of body centered particles was heated up and equilib-

rated at a temperature of $T = 2.0$, which is well above the transition melting temperature of around $T \approx 0.14$. In order to generate the glassy samples, the liquified system had to be cooled down. A baseline configuration was made by cooling the sample down to $T = 1.0$. The reason for defining this as a baseline configuration, is that all further samples was prepared by directly cooling the sample at $T = 1.0$, in order to account for the effect of aging.

Aging is an important effect in so-called out-of-equilibrium systems like found for glassy dynamics. Aging, as the name implies, mean that both the structure and dynamics of a system can change as time progresses [44, 45]. Even though the relaxation time for a glassy system, by definition, exceeds any practical timescale, a gradual move towards an equilibrium is always present. Thus, small effects associated to aging is often observed after a quench of the system and must therefore be accounted for. In this work, all samples with temperature lower than the baseline configuration was made by quenching the sample at $T = 1.0$ and letting them equilibrate through a fixed number timesteps of one million.

3.3 Dynamic Heterogeneity analysis

3.3.1 Defining mobile particles

A simple way to investigate dynamic heterogeneity of the system is to consider displacements of the particles over a given time interval. The distribution of the displacement magnitudes of the particles is given by the self-part of the Van Hove function described in section 2.3. For the homogeneous system, this distribution is expected to be Gaussian with a mean described through the diffusion coefficient at the given temperature. Studies have however shown that the distribution becomes non-Gaussian as the system transitions into a glassy state [46]. The distribution adopts a large tail, which grows longer as the temperature is lowered. This indicates that some particles move a much greater distance than the bulk of the particles. These are the particles which will be referenced to as mobile particles. A straightforward method for differentiating between these mobile and in-mobile particles is to compare the non-Gaussian Van Hove function to a Gaussian reference distribution with the same mean and standard deviation. The cutoff displacement distance r^* can then be defined as the intersection point between the two distributions, thus isolating out the particles that moved further than expected, $r(t) \geq r^*(t)$.

The natural diffusion of particles makes the cutoff distance time-depended, which means that it needs to be evaluated at each intermediate time step. This is done by calculating a histogram for the displacements of the particles, and then using a quadratic B-spline interpolation to convert the histogram data into a continuous curve. An intersection point to the interpolated curve and a reference Gaussian curve is then numerically calculated to give the cutoff distance r^*. Figure 3.4 shows a sketch of the Van Hove function G_s, the linked Gaussian distribution G_0, and the cut-off distance r^*.

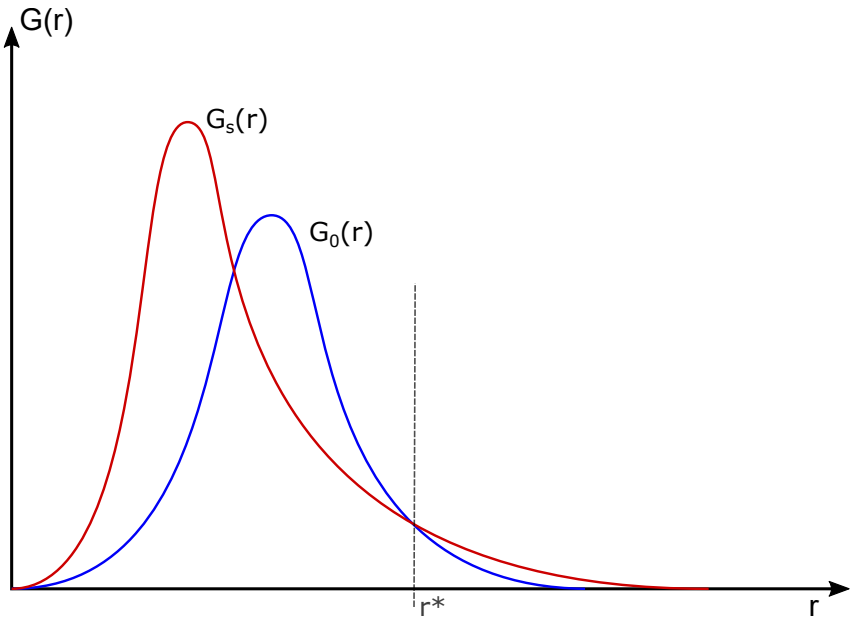


Figure 3.4: Sketch of the distribution functions describing the dislocation of the particles in the system. $G_s(r)$ represent the self-part of the Van Hove function, while $G_0(r)$ represent the Gaussian distribution with the same mean and standard deviation as $G_s(r)$.

3.3.2 K-means clustering

When the particles are split into mobile and in-mobile particles, heterogeneity in the particle positions is expected to occur for the glassy state in the caging region. This heterogeneity is visible through the forming of local clusters of the mobile particles. Quantifying this phenomenon can be difficult as the size and form of the clusters vary with time and they are often vaguely defined. One solution is to use a machine learning algorithm to assign the mobile particles into different clusters. One of the most common clustering algorithms is known as k-means. K-means produce a fixed number of clusters which all have a corresponding center. The number of clusters is usually given the symbol k, but to avoid confusion with the wave vector, it is here referenced by the symbol κ. All particles in the system are assigned to the cluster with the nearest center. The algorithm then adjust the cluster centers according to the particles assigned to each cluster. Next, all particles are again assigned to the nearest centers and the whole process is iterated until no new assignments occur. For a system with N particles the algorithm is designed to solve the optimization problem mathematically described as

$$\min_{\vec{\mu},z} \sum_{i=1}^{N} \| \vec{r}_i - \vec{\mu}_{z_i} \|^2, \tag{3.12}$$

where $\vec{\mu}_k$ is the center for the k-th cluster and z_i is the index for the cluster assigned to particle i.

The optimization problem is nonconvex and thus sensitive to the initial "guess" of cluster centers. Finding the optimal clustering is therefore not guaranteed for a given set of initial centers and one should choose this initialization carefully. Additionally, the worst-case convergence time for k-means is shown to be superpolynomial to the input size [47], which in this study can include up to thousand mobile particles. One of the most used methods for producing this initial "guess" in order to minimize the related problems, is a method called k-means++ [48]. The method starts by uniformly choosing a datapoint at random and letting this be the first cluster center. Next, the distance between all the remaining datapoints and the nearest cluster center, $D(\vec{r}_i)$, is calculated. The next center is then chosen at random from a weighted probability distribution where the datapoint $\vec{r}_i$ is chosen with probability proportional to $D(\vec{r}_i)^2$. These steps are then repeated until all κ clusters are chosen. By considering the squared distance to the nearest cluster center this method ensures a good spread of the centers, which is shown to improve the performance of k-means clustering.

Note that even with the improved method for finding the initial "guess" of the cluster centers, finding the global minima of the optimization problem is not guaranteed. In order to further ensure that a good clustering solution is reached, the k-means algorithm is applied to each configuration several times. The best solution of all iterations is then extracted according to the silhouette score, which is a metric that will be described later in this section.

One challenge with using k-means to find and quantify clusters of the mobile particles is that the algorithm requires the number of clusters κ as an input. This unknown for a system of particles and need to be estimated. There exist several good solutions for this problem, whereas many of them consists of applying the k-means algorithm on a range of values for κ. Then κ is chosen as the one which resulted in best clustering of the data. This method requires a metric for evaluating a cluster solution. The most used evaluation metric for this purpose is to look at the within-cluster-sum of squared errors (WSS). WSS is in fact what the objective function showed in (3.12) which k-means try to minimize. The method chooses the κ at which the WSS first start to diminish in the WSS-κ plot, a so-called elbow point. A problem occurs when the dataset consists of many non-distinctive clusters as this will make the elbow point to be vaguely defined.

A metric better suited for the clustering problem found in this study, is the silhouette score. The silhouette value for the i-th datapoint is given by

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}, \tag{3.13}$$

where a_i is defined as the average distance to the datapoints within the same cluster. b_i is defined as $\min_{k \neq z_i} b_{ik}$, where b_{ik} is the average distance from particle i to the particles in cluster k. By letting C_k be the set of particles assigned to the

cluster indexed by k, a_i and b_i can be defined by

$$a_i = \frac{1}{|C_{z_i}| - 1} \sum_{j \in C_{z_i}, j \neq i} \| \vec{r}_i - \vec{r}_j \| \tag{3.14}$$

$$b_i = \min_{k \neq z_i} \frac{1}{|C_{z_k}|} \sum_{j \in C_k} \| \vec{r}_i - \vec{r}_j \|, \tag{3.15}$$

where $|C_k|$ is the number of particles assigned to the cluster indexed by k.

Note that $-1 \leq s_i \leq 1$, where $s_i = 1$ indicate a perfect clustering of this particle as it lies well inside the assigned cluster compared to the surrounding clusters. Because of this, the average silhouette score $\langle s \rangle$ is a good metric for the clustering quality for a given solution. By plotting the average silhouette score as a function of cluster size, a maximum will occur at the optimal number of clusters for the specific dataset. More importantly, the value of this best average silhouette score itself is a good measure of how distinct the clusters appear in the system. Thus, for a system with particles uniformly and homogeneously distributed, the average silhouette score is expected to be low. An attempt to derive an analytic solution for the expected silhouette score was conducted, but was instead estimated through a Monte-Carlo process. For a system with N particles, the average of the top-scores from a significant number of samples with uniformly distributed particles were calculated and plotted against the number of clusters, κ, used in the k-means analysis.

A heterogeneous system where the particles group together is however expected to give a high positive value for the average silhouette score. Utilizing this principle gives a method for both approximating the best value for κ to use and evaluating the level of heterogeneity in a system. Because the solution given by the k-means algorithm is highly dependent on the initial "guess" of the cluster centers, a significant number of runs must be conducted for every configuration to ensure a good cluster solution. The method used to find the optimal κ for each timestep in a system, was to do a scan over a range of different values for κ and calculate the corresponding best silhouette score based on several runs with the k-means algorithm. Through a convergence analysis, the adequate number of runs at each value of κ can be estimated. Doing several thousands of runs for a wide range of values for κ can be a tedious process. To speed up the calculations, an initial coarse scan over values for κ is done. Here a smaller number of k-means runs are done to get an idea of where the optimal value can be. This narrows down the search area for the more precise scan where the adequate number of runs with the algorithm is done to find the optimal clustering solution and its corresponding silhouette score.

To further investigate the clustering of the mobile particles, the average cluster diameter for a solution can be calculated. A simple method for approximating the diameter of a set of points in space, is to calculate all the point-to-point distances within the set. The cluster diameter can then be set to be the maximum of these intra-cluster distances. Thus, or a the cluster indexed by k, the diameter d_k can

be defined as

$$d_k = \max_{i,j \in C_k} \| \vec{r}_i - \vec{r}_j \| . \tag{3.16}$$

Note that calculating all intra-cluster distances is a so-called brute force method. There exist much faster ways to estimate the diameter of a set of points, whereas many of these reduce the problem down to the convex hull of the point set. In this thesis, the number of particles assigned to each cluster was not immense, and thus calculating the cluster diameter proved to be fast with the brute force method. For systems of larger size, bigger fraction of mobile particles or bigger clusters, a faster algorithm to calculate the cluster diameters should be considered. An efficient alternative algorithm is described in [49].

Chapter 4

Results And Discussion

In this chapter, results from the static structure factor, intermediate scattering function and the heterogeneity analysis is presented. All results were calculated using trajectories from the molecular dynamics simulations as described in section 3.1 and appendix A, for the binary colloidal Yukawa model discussed in this thesis. In order to have a better statistical significance, a big number of data should be considered in a field like molecular dynamics, however, due to time limits, an average on ten samples maximum for each temperature was obtained and used for the calculations. Note that all measurements are given in reduced units, as mentioned in section 3.1.2.

4.1 The static structure factor

The static structure factor was calculated with trajectories of the binary Yukawa system in a glassy state at the MCT glass transition, i.e. $T = 0.14$. All particle type correlations were considered using the partial static structure factor given in equation (2.8) and averaged over a set of wave vectors with magnitudes in the shell $k + dk$. The resulting curves are shown in figure 4.1. For a finite simulation volume with periodic boundary conditions, the maximum possible period in the system is constrained by L. The cubic simulation box used in the simulations have side lengths of $L = 24.724$. Thus, any wave vector $\vec{k}$ shorter than $\Delta k = 2\pi/L \simeq 0.2187$ was not considered. The set of wave vectors used in the calculation was therefore chosen according to the following:

$$\vec{k} = (n_x \hat{e}_x, n_y \hat{e}_y, n_z \hat{e}_z)\Delta k, \tag{4.1}$$

$$n_x, n_y, n_z = \left\{ \mathbb{Z} \geq 0 \middle| \sqrt{n_x^2 + n_y^2 + n_z^2} \leq \frac{k_{max}}{\Delta k} \right\}, \tag{4.2}$$

where k_{max} is the maximum magnitude of $\vec{k}$ for which the static structure factor was calculated. Due to the direct calculation of the static structure factor being $\mathcal{O}(N^2)$ for each wave vector, this bound was set to $k_{max} = 15.0$ to reduce

the computational cost. More details on the optimization process for the static structure factor calculation is described in appendix B.

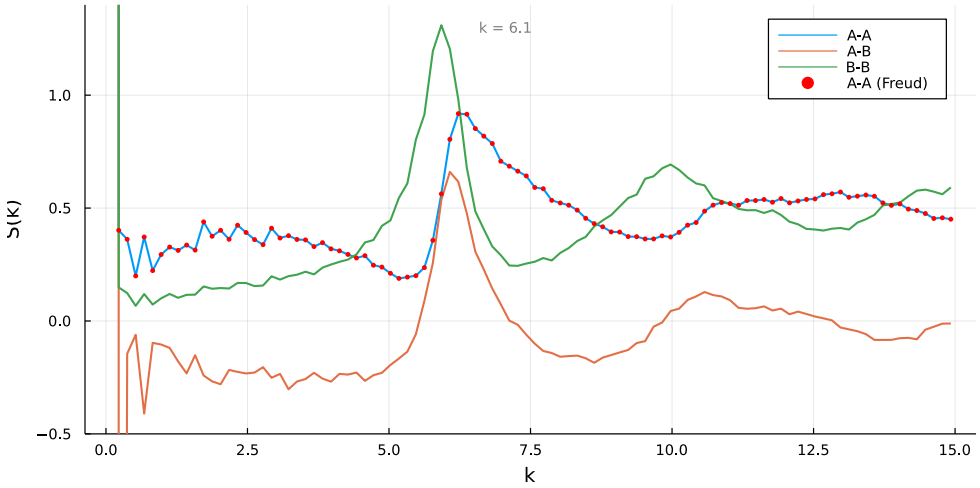


Figure 4.1: Static structure factor for the glassy-state Yukawa colloidal system at $T = 0.14$. The blue, orange and yellow lines represent the correlations for A-A, A-B and B-B particles respectively. The red dots represents a calculation of the A-A correlation obtained with the diffraction module of the freud Python library [50].

The static structure factor for A-A correlation calculated with the diffraction module of the freud.py Python package [50] was also plotted as red points in the same figure. As one can see, the results coincide perfectly with the one written in this thesis, this validates the codes for calculating the static structure factor. From the figure, it comes clear that the most significant wavenumber appears at $|\vec{k}| \approx 6$. Additionally, the correlation between the smallest type-A particles gives a main peak at slightly higher k value. This is expected, since to smaller particles it corresponds smaller periods, which result into higher wave vector because of the inverse proportionality to the period.

The sudden increase around $k = 0$ is a property of the static structure factor, $\lim_{k\to 0} S(k) = N$. A k-vector shorter than Δk is however not physical and thus the peak around $k = 0$ can be ignored. Also note that the plots of A-A and A-B is shifted by 0.5 downwards in order to better visualize all three curves. As the location of the peak, and not its actual value, is of importance, the downward shift will not have any effect on the following results. To conclude, $k = 6.1$ corresponds to the first Bragg peak in the static structure factor for our B-B correlations, and this number has been used as reference value to calculate the intermediate scattering function in the following parts.

4.2 The intermediate scattering function

The self part of the intermediate scattering function for the Yukawa binary colloidal system was calculated using equation (2.16). Figure 4.2 shows the resulting curves for temperatures ranging from $T = 2.0$ to $T = 0.005$. With a phase transition temperature around $T = 0.14$, this means that the figure include results from both liquid states and states transitioned into a glassy state. The magnitude of the wave vector was set to $k = 6.1$ as this corresponds to the main peak of the static structure factor.

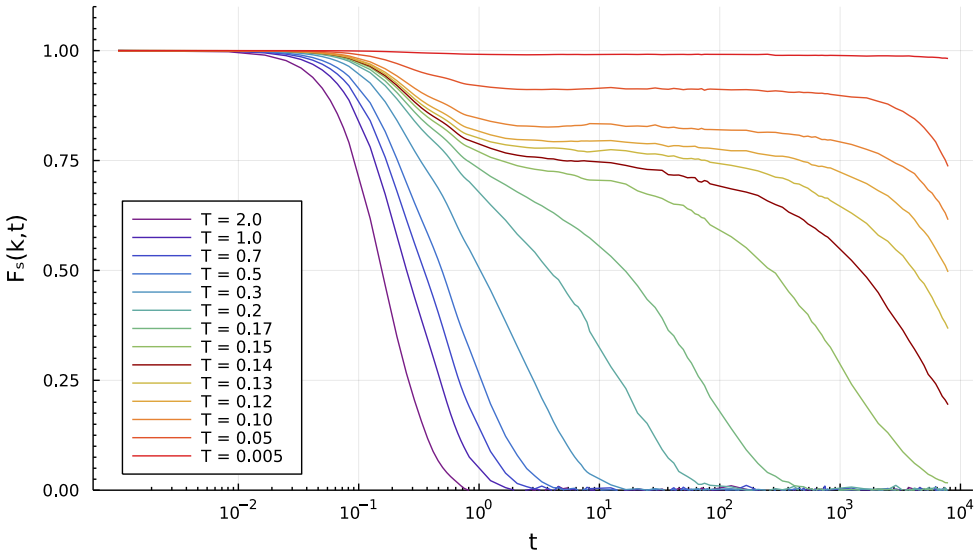


Figure 4.2: Intermediate scattering function of type-B particles in the Yukawa binary-colloidal system at different temperatures T. The magnitude of the wave vector was set to $|\vec{k}| = 6.1$ as this correspond to the main peak of the static structure factor.

At high temperatures, the system is in a liquid state and thus the particles are highly diffusive. As expected, the intermediate scattering function decrease down to zero after the system has passed through the initial ballistic region. Because of the high diffusivity in this liquid state, the spatial correlation to the initial configuration is lost early as the configuration changes fast. With a decrease in temperature for the system, this sudden drop in the intermediate scattering function appears to occur later. The delay of the drop can be explained by the temperature dependence of the diffusivity. With a lower temperature, the particles move slower and the correlation to the initial system is preserved for longer, i.e. a longer relaxation time.

At around $T = 0.2$, the dynamics however start to change. The sudden drop looks to be less steep and, as the temperature is lowered towards the phase transition, around $T = 0.14$, the plateau described in section 2.3 is starting to form. This plateau region is a result of the cage-effect which is characteristic for glassy

systems. The cage-effect means that most of the particles are trapped between slow and immobile particles, unable to break most of the correlation to the initial configuration. For systems around the phase transition temperature and lower, this plateau region is clearly visible and appearing with higher values for F_s. The duration of this period also looks to increase as the temperature is decreased and the period is a measure of the so-called β-relaxation. For the system used in this experiment, this β-relaxation time persists between the time 1×10^3 and 1×10^4, depending on the system temperature. Thus, with a lower temperature, particles of the glassy system tend to preserve the correlation, or "memory", of the initial system for longer as slower particles will use longer time to escape the cage-effect.

At even lower temperatures the particles of the system are almost completely frozen, and the configuration will not change considerable until large timescales are met. Because of this, F_s is held around one and the plateau is not evident.

The α-relaxation is related to the timescale at which the intermediate scattering function breaks down to zero and all "memory" of the initial configuration is lost. For liquid states, the relaxation times is easy to quantify because of its rapid breakdown. The relaxation times is however harder to quantify for the glassy states because of the low diffusivity and caging-effects. Mode coupling theory predicts curves for both the α- and β-relaxations, as described in section 2.4, with the stretched exponential. The curves of the intermediate scattering function results for the Yukawa colloidal system were fitted using equation 2.17. The fitted curves are presented as dashed lines in figure 4.3, with their respective fitted parameters for β and τ.

For the liquid system at $T = 1.0$, only the α-relaxation is relevant. The fitting function was optimized to fit the drop of $F_s(k, t)$ in order to predict this relaxation process. The curve looks to fit the intermediate scattering function up to the inflection point with the parameters $\beta = 1.7$ and $\tau = 0.29$, where the latter reflect the relaxation time. This is what is expected for a liquid state, i.e. the parameters are consistent with the typical time-decay Debye-relaxation described in Section 2.3, consistent with many experiments and previous simulations on very similar colloidal systems [2, 4, 51] For $T = 0.14$, the dynamics of the glassy states delays the relaxation. The drop in $F_s(k, t)$, subsequent to the ballistic regime, now represent the so-called β-relaxation. However, the parameters from the fitting present a higher value for τ, how it is expected for a cage-type region characterized by a slowdown in the dynamics. The final relaxation, the α-relaxation, is also predicted with the typical stretched exponential as a fitting curve that validates a typical glassy, non Gaussian behavior as described in previous papers for similar systems [2, 51]. The exponent of the time-decay is estimated to be $\beta = 0.62$, which agree with predictions from mode coupling theory stating $\beta < 1.0$. Further, the relaxation time is here associated with $\tau \approx 10\,640$. In the paper form Kikuchi et al. [51] a colloidal binary mixture modeled with Yukawa potential is investigated. The binary mixture is very similar to the one investigated here, and one can observe similar trend in the self-intermediate scattering function. This agreement further validates the investigations.

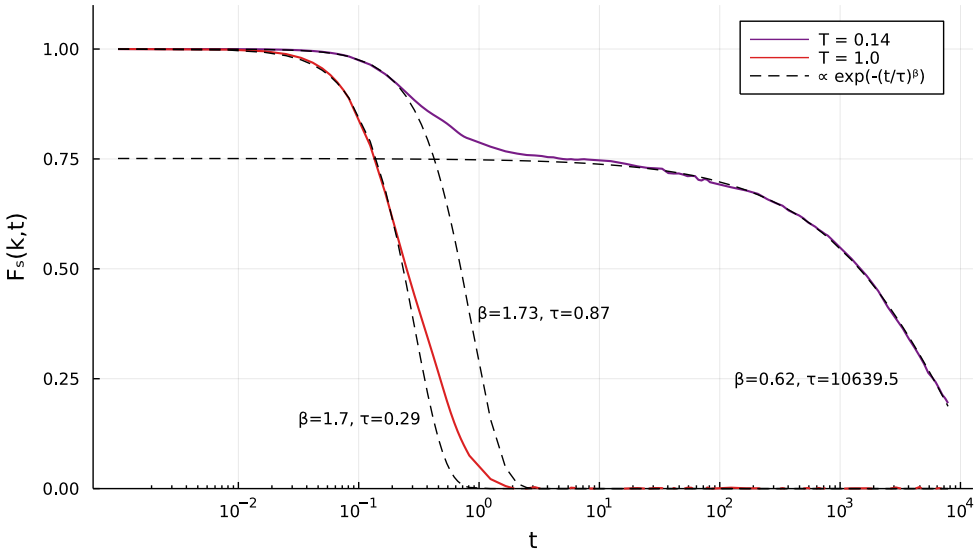


Figure 4.3: Intermediate scattering function of type-B particles in the Yukawa binary-colloidal system for a liquid and a glassy state. The magnitude of the wave vector was set to $|\vec{k}| = 6.1$ as this correspond to the main peak of the static structure factor. The dashed line represents fitted curves with forms predicted by the mode coupling theory.

4.3 Dynamic heterogeneity

In this section, the traditional way of studying the dynamic heterogeneity is first presented by looking at the non-Gaussian parameter and the dynamic susceptibility. Then the mobile particles are extracted before the method of k-means is finally applied and analyzed.

4.3.1 α_2-parameter and dynamic susceptibility

In order to study the dynamic heterogeneities of the glassy system, one first must get an idea to where the heterogeneities are occurring. Figure 4.4 illustrates this by plotting the non-Gaussian parameter $\alpha_2(t)$ for the Yukawa colloidal system at different temperatures according to equation (2.18). Each line is an average of ten samples with an additional simple moving average applied to highlight the trends of these curves. The two liquid systems look to remain at zero during the whole simulation. This is expected for a system with particles not bounded by its configuration and thus its displacement distribution will take a Gaussian form. The same cannot be said for the glassy systems transitioned out of the ballistic regime. Here $\alpha_2(t)$ grows, which is expected to be related to the cage-effect holding parts of the configuration locked, and thus the development of a tail in the Van Hove function appears. Further, the peak of the non-Gaussian parameter increase as the temperature is lowered, which is exactly the same trend found in previous

work discussed in section 2.5. The decrease of temperature also predicted a time-shift for this peak, and the same trend is found here. The delay of the peak for lower temperatures relates to the higher relaxation times for these systems. By comparing these results with the intermediate scattering function, one can also note that $\alpha_2(t)$ does not start to grow before the plateau in $F_s(k,t)$ appear.

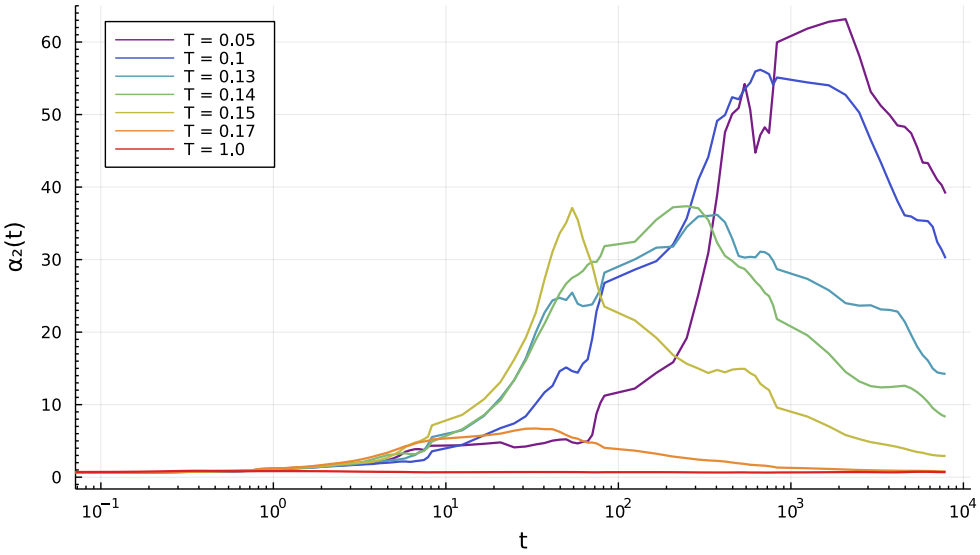


Figure 4.4: The non-Gaussian parameter for the Yukawa colloidal system at different temperatures. All lines represent an average of ten samples for the given temperature. Additionally, a simple moving average was applied on the timeseries.

In figure 4.5 the susceptibility $\chi_4(t)$ of the system is plotted for the Yukawa colloidal system at different temperatures using equation 2.19. Note that the variance of the fluctuations is calculated using the simulations of ten different samples for each temperature. As $\chi_4(t)$ is not an ensemble average like many other measurements in this thesis, it requires a higher number of simulations to ensure accuracy. Limitations in time deprioritized this issue, and ten samples was found to be adequate to the qualitative analysis of this measurement. A simple moving average was applied to smooth out the curves.

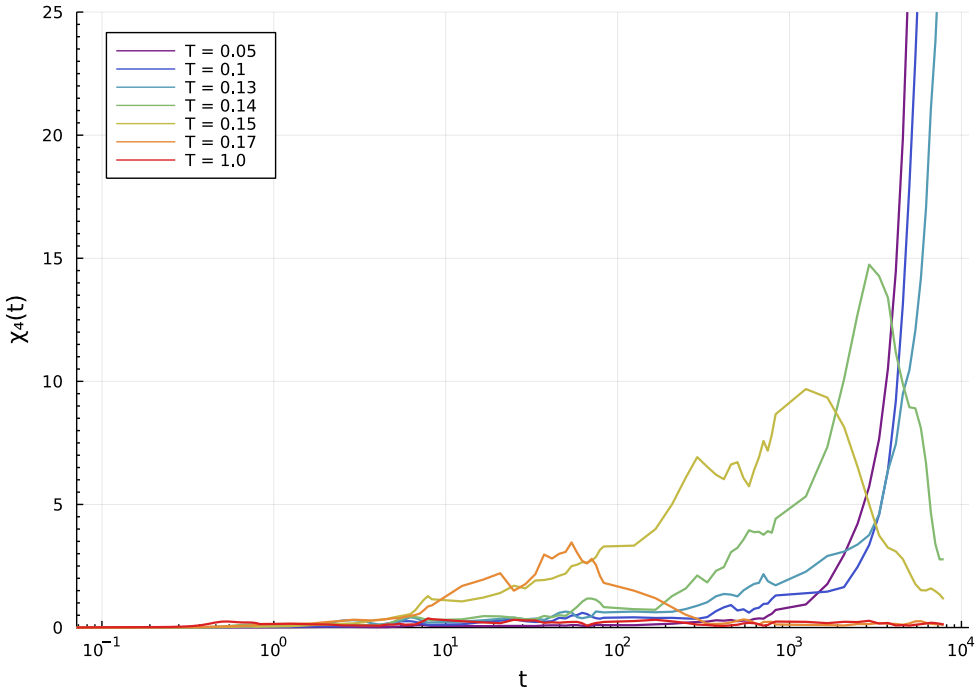
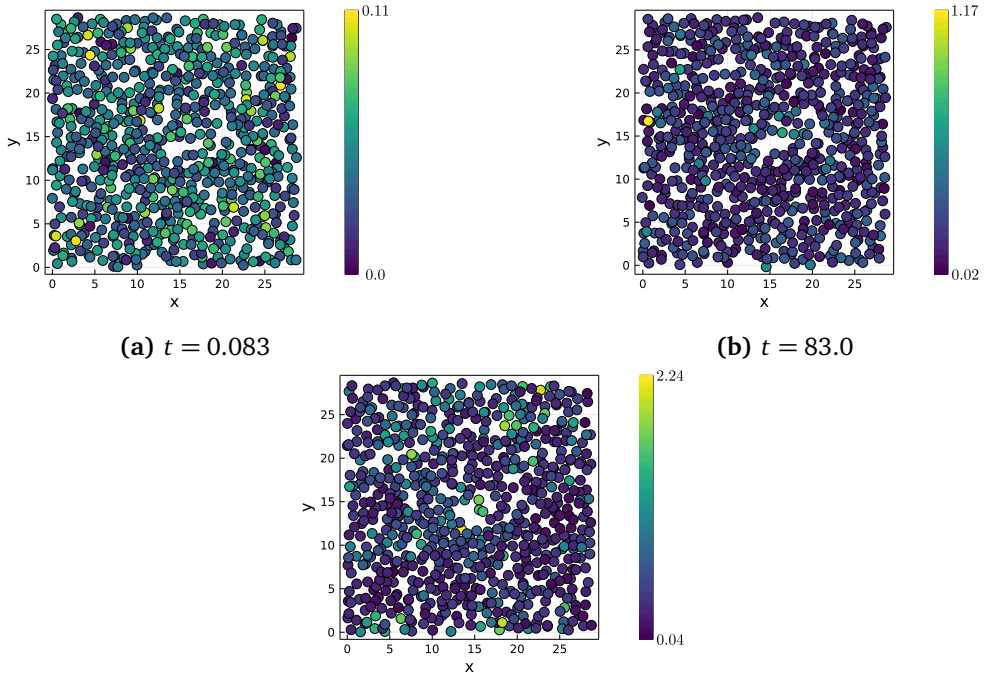


Figure 4.5: The dynamic susceptibility for the Yukawa colloidal system at different temperatures. All lines represent an average of ten samples for the given temperature. Additionally, a simple moving average was applied on the timeseries.

Figure 4.5 show clearly three peaks at T=0.17, 0.15 and 0.14. The peaks of the susceptibility for lower temperatures T $<$ 0.14 appear to occur at later times and thus not caught in the timespan of the calculations used in this work. However, the expected trend is evident with the growth of the peak as the temperature is decreased below the transition temperature. The overall shape of the curves, especially as seen for $T = 0.14$, coincide with results of many other simulation works in which glassy states and DH are studied [6, 9].The trend is as that for a decrease of temperature, both, the $\chi_4(t)$ and $\alpha_2(t)$ increases in peak height and their peaks shift to larger times [19].

4.3.2 The mobile particles

Figure 4.6 shows a snapshot in the plane $z = 10.0$ of the Yukawa colloidal system for type-B particles at $T = 0.14$. Each subfigure illustrates the different regimes that takes place in the simulation, namely the ballistic-, cage- and diffusive regime. The colors are mapped according to how much each particle are dislocated in the given timeframe. One can clearly see how the highly dislocated, mobile, particles seem to cluster together in the later times, subfigure 4.6c, especially compared to the initial ballistic regime illustrated for $t = 0.083$, subfigure 4.6a, and cage-effect time, i.e. time $t = 83$, subfigure 4.6b.



(a) $t = 0.083$

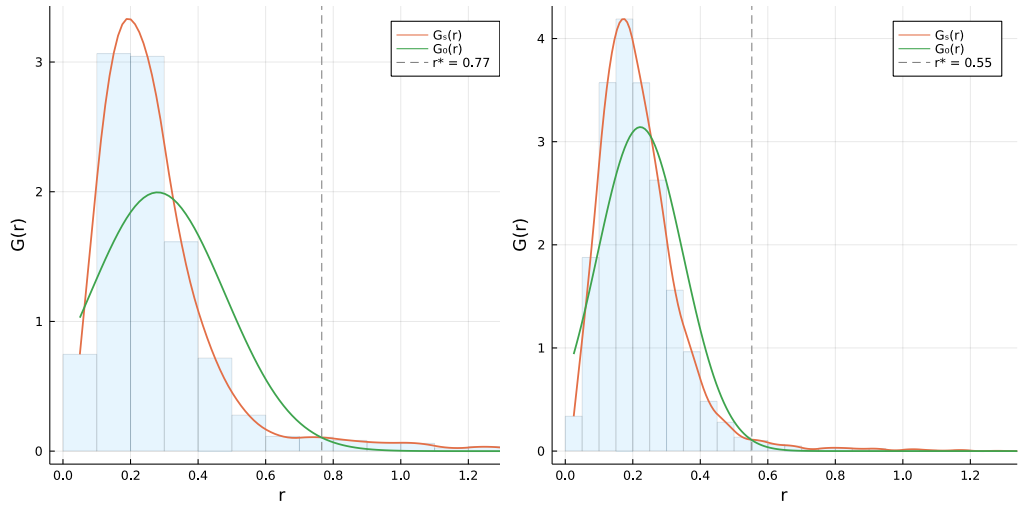
(b) $t = 83.0$

(c) $t = 2490.0$

Figure 4.6: Snapshots in the plane $z = 10$ of the Yukawa colloidal system at $T = 0.14$ for three different timesteps. The colors are mapped according to how much each particle have dislocated. Note that only type-B particles are plotted, and size is not to scale.

In order to perform the heterogeneity analysis on the mobile particles in the system, they first need to be defined. For each timestep, the Van Hove function was calculated and compared to the Gaussian distribution with same mean and standard deviation, as described in section 3.3.1. Figure 4.7a and 4.7b shows the Van Hove function and its corresponding Gaussian distribution for the glassy state with $T = 0.14$ at time $t = 84.0$ for type-A and type-B particles respectively. One can notice that the Van Hove function appear to develop a tail indicating that few particles displace further than the bulk of the remaining particles. These particles are defined as the 'mobile particles' of the system. As described earlier, this tail is related to that the bulk of the particles are still spatially correlated and "caged". Only a fraction of the particles are expected to break free into a dislocation that form the tail of the Van Hove. Using the crossing point, r^*, between the Van Hove and the Gaussian appear to be a good method for extracting this tail, and thus differentiate between mobile and immobile particles.

Although the method for extracting the mobile particles seem to work for the glassy state at $T = 0.14$, its soundness depends on the Van Hove function having a distinct tail with dislocated particles. For the liquid state and glassy states in the ballistic regime, this tail however showed to be less distinct, which is due to the particles being less bounded in a specific configuration and free to diffuse.



(a) type-A **(b)** type-B

Figure 4.7: Calculation of cutoff distance r^* using the Van Hove function $G_s(r)$ and a Gaussian distribution $G_0(r)$ for a system at temperature $T = 0.14$ after a time $t = 83.0$. Results for type-A particles (a) give an intersection point at $r^* = 0.77$, while type-B (b) gives an intersection point at $r^* = 0.55$.

Thus, the spatial correlation is not conserved for these systems, and it makes the Van Hove function Gaussian for the liquid states, as one should expect. This was confirmed by the low non-Gaussian parameter found for the liwuid system.

In order to keep the Van Hove definition consistent, a fixed number of bins was used. Through some trail and error, the number of bins was set to 25, as this gave reasonable curves for the whole range of timesteps and temperatures investigated in this thesis. A bigger number of bins resulted in an unsmooth curve which made the crossing point of the Gaussian and the Van Hove ambiguous. And furthermore, a smaller number of bins resulted in a too coarse description of the Van Hove. Note that the two plots in figure 4.7 seem to incorporate a different number of bins in the Van Hove, but this is due to the type-A particles being of smaller size. The smaller diameter of type-A particles speed up the dynamics compared to the larger particles, and thus the tail in the Van Hove is developed earlier. With a longer tail and a fixed number of bins, the size of the bins is simply increased. It can be assumed that an adaptive number of bins would have made the definition of the Van Hove function more accurate, but this was not investigated.

Figure 4.8 shows the development of the cut-off distance r^* as function of time illustrated in a log-log plot for three different temperatures. Each line is given by the average over five different samples. Notice how similar the curves look to what is expected for the motion of the particles in this regime. As the cut-off distance r^* is highly dependent on the displacements of the particles, the curves can be assumed to represent the motion of the particles.

The initial increase in r^* represents the ballistic regime and as expected, the three different temperatures behave similar here as none of the particles have had time to be affected by their surroundings. The linear increase shown in the log-log plot for r^* in this region was found to be around $c_1 = 1$. Keep in mind that the the mean squared displacement of the particles is found to be of second order in the ballistic regime. Thus, an initial first order increase of the cut-off distance should be no surprise if one assumes that it is linearly dependent on the mean displacement of the particles [32].

For intermediate times the liquid state, $T = 1.0$, continues to grow further, but with a slightly lower rate of $c_2 = 0.51$, which was found through a linear regression. The decrease of the slope indicates that even for liquid states, the interactions between the particles give a lower diffusion of the particles. Similarly, this behavior is also found in the analysis of the motion for the same system. This region is the diffusive region and the transitioning from the ballistic regime is given by the intersection of the two regression lines. The transition point was estimated to be at $t = 0.29$. If one compares this result to the fitting curves on the intermediate scattering function from figure 4.3, the point appear to occur exactly at the predicted α-relaxation time for the system at $T = 1.0$. This shows that the cut-off distance used to characterize the mobile particles incorporate the dynamics of the different regimes, which further improves the validity of this method.

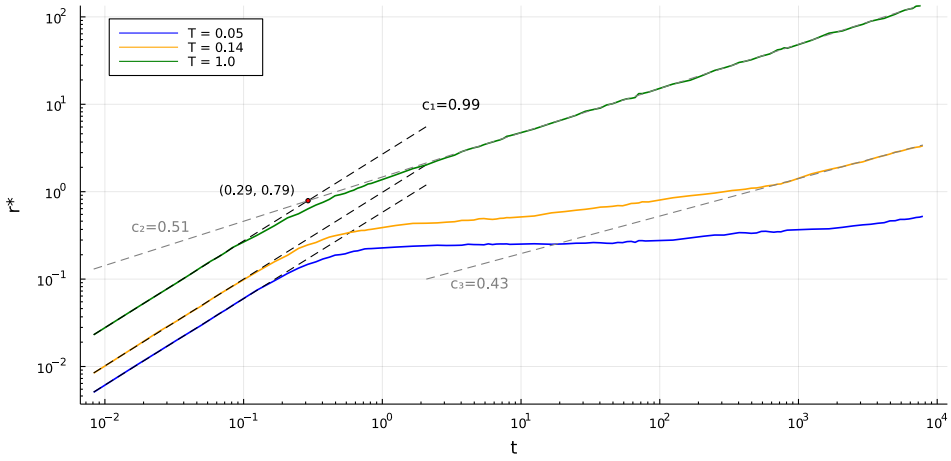


Figure 4.8: Log-log plot of defined cutoff displacement r^* for type-B particles in liquid state, $T = 1.0$, and two glassy states, $T = 0.14$ and $T = 0.05$ as a function of time t. The dashed lines c_1, c_2 and c_3 represent the linear regressions of the ballistic regime, the diffusive regime for $T = 1.0$ and the diffusive regime for $T = 0.14$ respectively.

For the glassy state at $T = 0.14$, the growth of the cut-off distance halts after the initial ballistic regime. This transition into a nearly fixed cut-off distance can be explained by the cage-effect discussed earlier in which only a small fraction of the particles is expected to dislocate in any significant amount. Also here the

connections to the dynamics found in the intermediate scattering function can be drawn. The cage regime locks the configurations until the system eventually transition into a diffusive region at higher timescales. Here many of the particles are able to break the caging and most of its spatial correlation starts to fade, which correspond to the dynamics found in the intermediate scattering function. Notice how the time at which r^* starts its second growth agree with the time at which the correlation described through the intermediate scattering function break down. The slope at the diffusive region was estimated through a regression which gave $c_3 = 0.43$, although the growth in this region is expected to be the same as for the liquid state in the diffusive region. The slightly lower growth in the diffusive region compared to $T = 1.0$ can be explained by that the diffusive region for $T = 0.14$ was not fully achieved in the limited timescale used in these simulations.

For the system at $T = 0.05$, the caging region is much more evident as the growth of the cut-off distance appears to be zero after the transition from the ballistic regime.

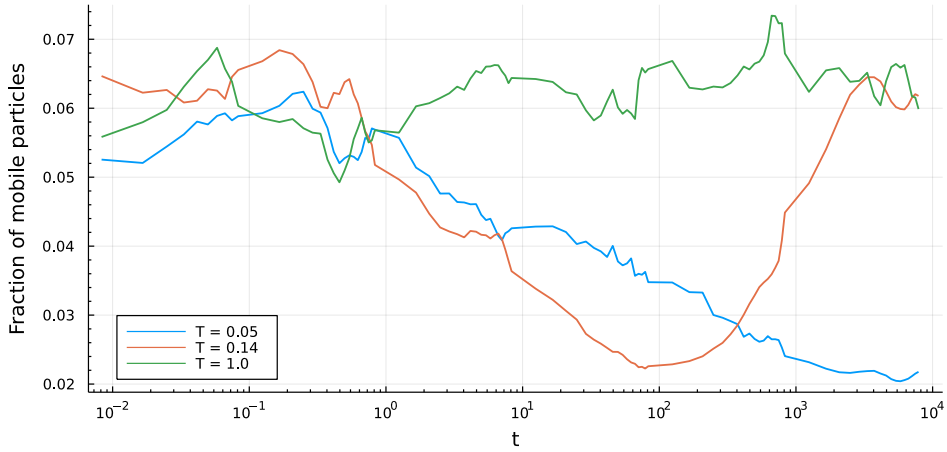


Figure 4.9: Semi log-plot of the fraction of mobile type-B particles in the Yukawa colloidal system simulated for three different temperatures. Each line represent an average over five different samples at the given temperature and a simple moving average is applied.

Figure 4.9 shows another interesting result form the mobile particle calculation. Here the fraction of mobile type-B particles is plotted by dividing the number of mobile particles for each timestep by the total number of type-B particles in the system, which in this case was 8000. Each line represent an average of five different samples, and a simple moving average of five datapoints was further used to smooth out the plots. The liquid system at $T = 1.0$ does not seem to present any significant temporal change during the timeframe of this experiment and only fluctuate around a fraction of 6% of the total particle count. This is not a surprise

considering that the liquid state is not bounded by its configuration through the cage effect, and this has also been declared by other previous works [11]. The Van Hove function is thus more similar to a Gaussian distribution with no forming of a tail like in the glassy states. As a result, the dynamics of the system is only changing in the transition from the ballistic to the diffusive regime. This transition is however not visible in this calculation. An explanation of this could be that the form of the Van Hove function is conserved during this transition, and the change of the dynamics is corrected through the change of the cut-off distance shown in figure 4.8.

In the glassy states at $T = 0.05$ and $T = 0.14$ the three different regimes is more distinct. As expected, and what we have seen in previous calculations, they all share the same characteristics in the ballistic regime [9]. In the intermediate timescales, the glassy states however adapt a gradually smaller fraction of mobile particles after they transit into the cage regime. This lower fraction of particles can be explained by that the tail of the Van Hove function is gradually increasing as more and more particles are "escaping" their cage of neighboring particles [11]. The growing tail increase the relative size of the standard deviation for the Van Hove function, which again result in a wider Gaussian distribution. With the method used to extract the mobile particles, a wider Gaussian distribution effectively increases the requirements for a particle existing in the tail, and thus the number of mobile particles decrease.

The decrease in the fraction of mobile particles can further be explained by the development of heterogeneity in the system. As described, the Van Hove function acquires a more distinctive tail as the glassy systems transition into the cage regime. The fact is that this tail is a result of the heterogeneity, and one can not really talk about a significant tail in either the ballistic regime, or the liquid state. Thus, the systems at $T = 0.14$ and $T = 0.05$ experience a development of heterogeneity when the mobile particles describe the particles that "escape" their cage rather than only the most diffusive particles.

One should also note that the glassy system at $T = 0.14$ returns to the fraction of mobile particles found in the liquid state as the system enters the diffusive region. The timescale of this return correlates with the transition time which was found in the calculation of the cut-off distance. Similarly, the glassy system at a much colder temperature at $T = 0.05$ is expected to jump back to a mobile fraction of 6% as the diffusive region is reached. The delay of this transition is explained through the slowdown of the dynamics as the temperature is lowered

In figure 4.10 a snapshot of the Yukawa colloidal system at $t = 83.0$ with $T = 0.14$ is plotted. The two different types of particles are represented by the two colors and only the mobile particles are visualized. For the given time and temperature, one can see from the earlier analysis that the system should experience a degree of heterogeneity. In the figure, the particles look to be clustered together. Note that these clusters seem to be independent of particle type. An explanation for this clustering could be described by the dynamics forming under the cage-effect. With most of the particles trapped by the neighboring particles, a

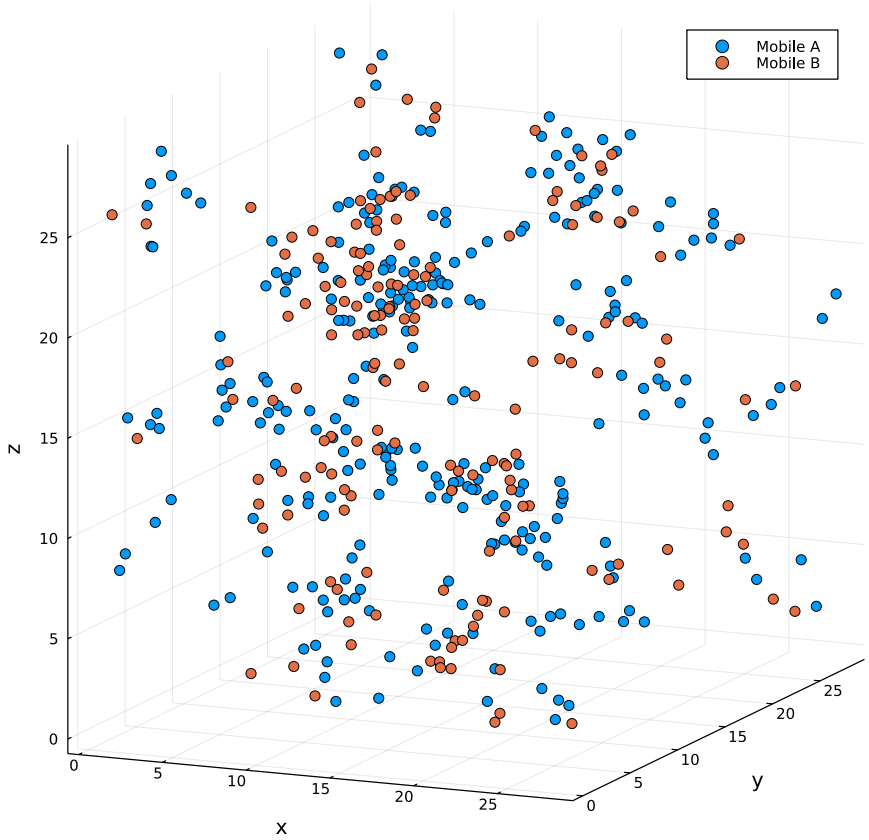


Figure 4.10: Snapshot of the mobile particles of the Yukawa colloidal system in a glassy state, $T = 0.14$, at a time $t = 83.0$. Blue points represent type-A particles, orange points represent type-B particles, and their size is not to scale.

sudden escape for one of these particles could resolve some of the tension between the particles. This change of the local configuration could trigger the surrounding particles to dislocate into new positions. With this understanding, the mobile particles should form in groups like we see in the figure. Many papers have also spoken of mobile particles clustering into "string-like" shaped clusters, contrarily to immobile particles that assume a more spherical appearance [9]. One can see the same behavior in here, the correlated mobile particles are gathering in string-like shaped clusters.

4.3.3 Clustering analysis

With the mobile particles now extracted, the clustering algorithm can finally be used to quantify the heterogeneity of the system. To ensure accurate measurements using the k-means algorithm, a convergence analysis was performed and described in Appendix C. The main finding from the analysis is that convergence

required a bigger number of iterations than first anticipated. The qualitative results however showed similar conclusions even when the number of iterations were not adequate to ensure convergence.

Because the inherent number of clusters, κ, in each configuration was unknown and difficult to distinguish with a standard approach, The method described in section 3.3.2 was decided as a method to find the optimal values for κ. The scan over κ used in the k-mean algorithm ranged from two clusters to 50% of the total number of mobile particles with an upper limit of 150 clusters. This ensured that each cluster included at least two particles at average. The silhouette scores in the first scan were calculated as the best score among 30 iterations for each κ in the scan. The second and more precise scan utilized 500 iterations for each κ that ranged over the 21 κ-values surrounding the peak of the first scan. The exact code used in this calculation is described in Appendix D. All of the following clustering results were calculated with k-means described in the Chapter 3.

A result of the applied k-means algorithm is shown in figure 4.11. The figure visualizes the mobile particles of the Yukawa colloidal system at a temperature of $T = 0.14$ for $t = 83.0$, which is the exact same snapshot as in figure 4.10. The color coding of the particles in figure 4.11 was given to represent the clustering solution which resulted in the highest silhouette score. The highest score was calculated to be $\langle s \rangle = 0.451$ and given with the number of clusters set to $\kappa = 32$. Thus, each of the 32 clusters is given its own color, although it can be somewhat difficult to differentiate them in the center of the figure. For the clusters positioned along the edge of the system, the clustering solution clearly seem to be good at defining the borders between each cluster.

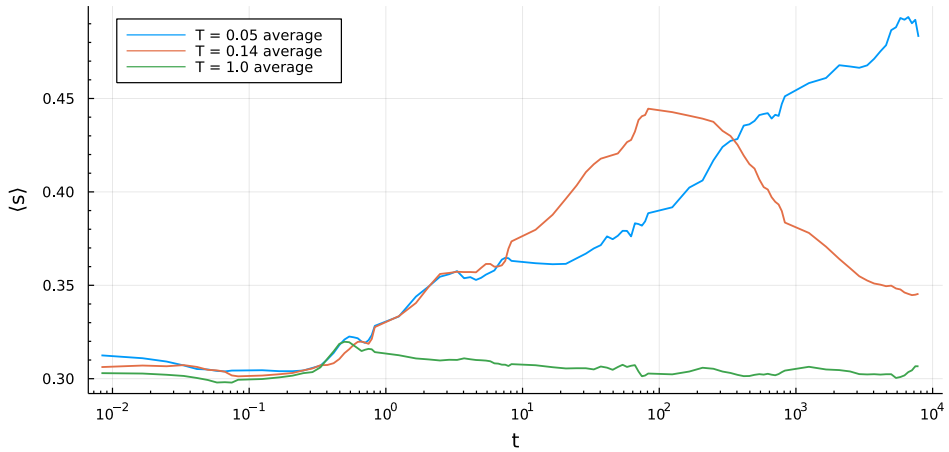


Figure 4.12: Average silhouette score $\langle s \rangle$ for the mobile type-B particles in the Yukawa colloidal system at three different temperatures. Each of the lines represent an average over five different samples at the given temperature.

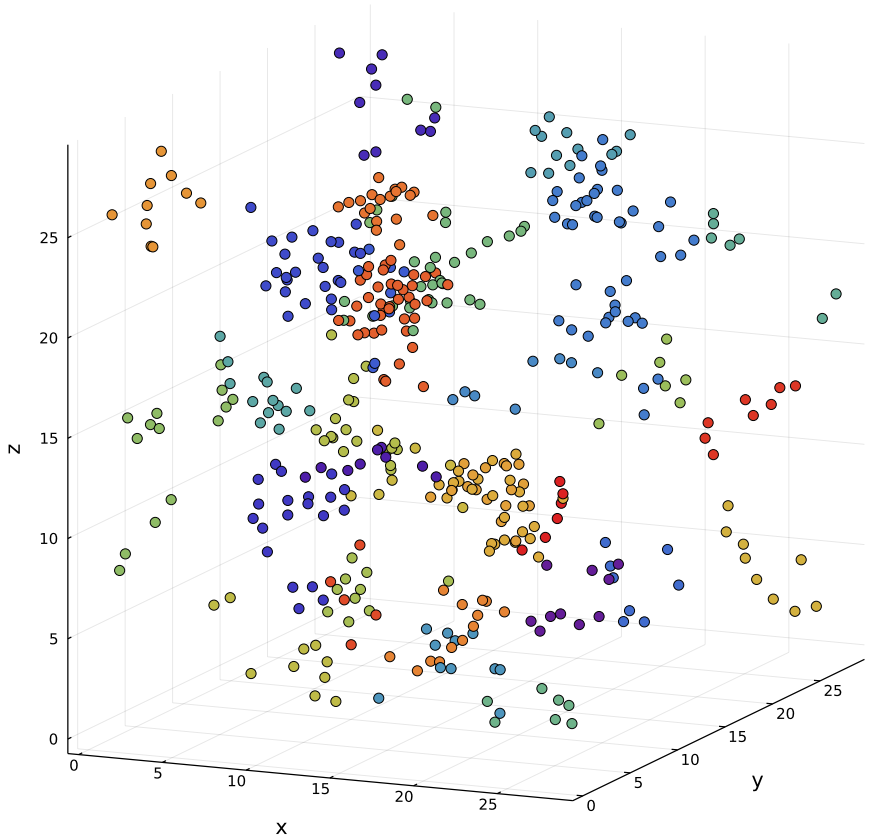


Figure 4.11: Snapshot of the mobile particles of the Yukawa colloidal system in a glassy state, $T = 0.14$, at a time $t = 83.0$. The k-means algorithm is applied to cluster the particles and each color represents a specific cluster. The number of clusters used in the calculation was $k = 32$, which gave an average silhouette score of $\langle s \rangle = 0.451$. Size of particles are not to scale.

Figure 4.12 shows time dependence of the calculated average silhouette score for the mobile type-B particles in the Yukawa colloidal system simulated at different temperatures. Each line represents an average over ten different samples with a simple moving average including five data-points applied to smooth the curves without losing qualitative information on the dynamics.

As expected, both the liquid and glassy states have similar scores in the ballistic regime. As explained earlier, the particles in this timescale do not have sufficient time to affect each other, resulting in a homogeneous blend in the dislocations for the various particles. The mobile particles should thus be scattered around in no particular order and the fact that the glassy state show similar values for $\langle s \rangle$ as the liquid state, indicates that no heterogeneity is present in this region. The transition out of the ballistic regime marks an increase in $\langle s \rangle$ for all temperatures. One explanation for this relatively small increase during the transition can be that

the transition does not occur in a uniform fashion. Local variation in temperature across the simulation volume can result in some regions of particles transitioning faster than others. Thus, with the definition of mobile particles used in this thesis, heterogeneity will occur when only these transitioned particles are filtered out. This explanation requires that a homogeneous system like the liquid state should return to a lower silhouette score after the transition into a diffusive regime, which it in fact does for the rest of the simulation time.

For the two glassy states, the increase of $\langle s \rangle$ however continues well after the transition from the ballistic regime. As seen in the intermediate scattering function for the system, they transition into the cage regime, and it is exactly here the heterogeneity of the glassy states are predicted to emerge. In logarithmic time, a seemingly linear increase of the silhouette scores is present with the same rate for both temperatures. For the system at $T = 0.14$, this behavior change at times in the order of 1×10^2, which not surprisingly also marks a transition of the dynamics seen in the intermediate scattering function. Here the particles in the system starts to migrate into a diffusive region and relaxing the system. A peak of $\langle s \rangle \approx 0.44$ is a decent silhouette score and can only be explained by a sufficient heterogeneity in the configuration, with evident clusters. For longer times, this heterogeneity should fade out as the system relaxes and the silhouette score is expected to idle down to a value around 0.3. The exact same behavior is visible for the glassy system at a colder temperature of $T = 0.05$. One of the differences at this temperature is that the almost "frozen" particles result in a much slower dynamic, which again give a delayed peak in the silhouette score and the transition to a diffusive regime. A more interesting difference is the higher peak value of the silhouette score. A possible explanation for this is discussed later.

The silhouette scores calculated from the system at different temperatures can be compared with respect to each other. The results alone can however not support a nonzero value for $\langle s \rangle$ at configurations with mobile particles predicted to be homogeneous in space. To further explain the results found from the clustering analysis, a truly homogeneous system was calculated by using k-means on a system of uniformly distributed particles. The average silhouette score is plotted in figure 4.13 as a function of κ for systems containing $N = 100$, $N = 500$ and $N = 2000$ particles positioned in a fixed volume equal to the simulation box used to simulate the Yukawa colloidal system. The results show an exponentially decreasing silhouette score as the number of clusters increases, before it eventually goes to zero when the number of clusters is equal to the number of particles in the system. When κ is equal to the the number of particles in the problem, each particle will act as a separate cluster. From equation (3.13) one can see that this makes $a_i = 0$, giving $s_i = 0$. Thus, the highest silhouette score for a uniform distribution is given for the lowest value of κ. In this thesis, the fraction of mobile particles ranged from 2% to 7% from a total of 8000 type-B particles, meaning that an average silhouette score between 0.30 and 0.35 is expected for a homogeneous configuration. This coincides well with the results presented in figure 4.12, for the systems in the ballistic regime, and the liquid state. Furthermore,

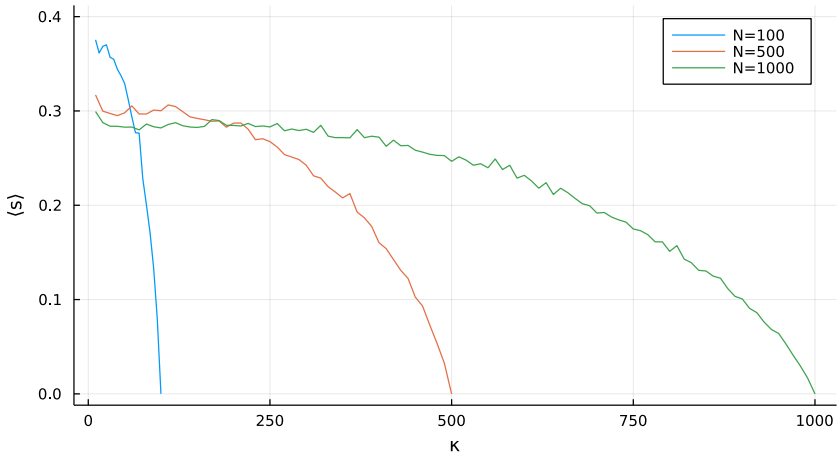


Figure 4.13: Average silhouette scores calculated as a function of number of clusters in a system, κ, of uniformly distributed particles positioned in a volume corresponding to the simulation box of molecular dynamics simulations. The scores were calculated for systems with particle numbers of $N = 100$, $N = 500$ and $N = 2000$.

these results show that the silhouette score is dependent on the particle density in the configuration, and thus the fraction of mobile particles should be considered when discussing the silhouette scores of the Yukawa colloidal system. At the same time, the advantage of having a lower density is far from explaining the high peaks of the silhouette scores found for the glassy system.

In addition to the silhouette scores, other interesting results could be extracted from the clustering solution given by k-means for each configuration of the Yukawa colloidal system. Figure 4.14 how the average number of particles for each cluster is changing over the timescale of the molecular dynamic simulation for three different temperatures. Further the average cluster diameter was also calculated as described in section 3.3.2, and the results are shown in figure 4.15. Equivalent to earlier results, the different regimes in the dynamics are clearly present in these results. The ballistic regime shows that the best silhouette scores were given for big clusters, which coincide with the results found for the system of uniformly distributed particles. At the transition from the ballistic regime, true clusters are believed to form for all temperatures. Thus, the optimal cluster solution is suddenly given by the inherent number of clusters, which in this case give small clusters each containing only a few particles. For further timescales, the dynamics of the liquid and the glassy states separate as the liquid state gradually fall back to big clusters. As expected for the glassy states, the trend of small distinct clusters continue into the cage regime, while one can see slight increase in cluster size as the system at $T = 0.14$ move into a more diffusive regime at high timescales.

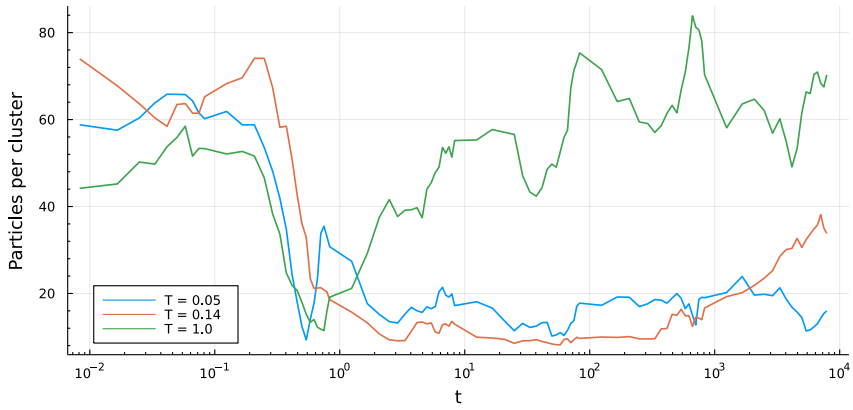


Figure 4.14: Average number of particles per cluster for the type-B particles in the Yukawa colloidal system at three different temperatures. Each of the lines represent an average over five different samples at the given temperature.

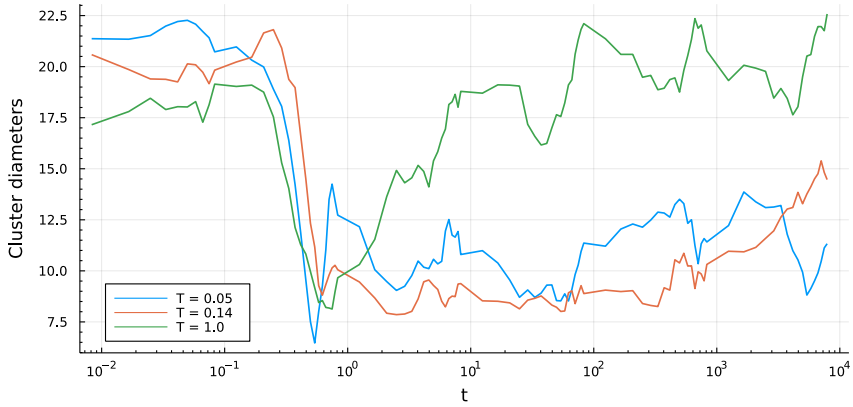


Figure 4.15: Average cluster diameter calculated for the mobile type-B particles in the Yukawa colloidal system at three different temperatures. Each of the lines represent an average over five different samples at the given temperature.

A somewhat surprising discovery is that the for the glassy states, both the fraction of mobile particles and the silhouette scores seem to evolve in an exponential fashion in the cage-regime, while the size of the clusters seem to be confined in time for the same region. When the glassy states incorporate the cage-effect, the mobile particles one by one hop out of their cages, resulting in a growing tail of the Van Hove. This increase of the tail will further highlight the difference between mobile and immobile particles. Thus, making it less likely that a relatively fast, but not fully mobile, particle will fall under the definition of mobile particles used in this thesis. So even though the true fraction of mobile particles increases, less of the bulk of particles are included in the definition, and the effective number of

mobile particles decrease. With a more well-defined group of mobile particles, the silhouette score increases accordingly, as it is these particles that should invoke heterogeneity.

The reason for the more conserved cluster size is however harder to consider. One explanation can be that the truly immobile particles included in the definition can simply fall both inside and outside the proximity of the true clusters. By doing an average over all the clusters, the effects cancel each other, keeping the size of the clusters connected to the heterogeneity of the mobile particles in the given configuration.

As a final note, it should be said that recently, many other ML approaches have been used successfully to study the glassy state DH in order to provide properties predictions [17, 18, 20, 21]. In general, all the ML approaches, included the one studied in this thesis, have been able to represents some features of the glass phenomenology. Thus, validating the meaning of those tools are ideal fitting methods.

Chapter 5

Conclusion

In this thesis, I have studied the relaxation and dynamical Heterogeneity of a Yukawa-glass forming material combining Molecular Dynamics (MD) simulation with the k-means clustering technique. During my pre-study work, I have got used with the MD technique and the simulation model, i.e. Yukawa binary colloidal system, that I have shown to present the typical feature of a glass forming material, reproducing with an excellent agreement previous results [16, 22, 51]. Being sure to have found the right model, I have decided to spend the months of my Master in studying the intriguing problem of the glassy dynamics and Dynamics Heterogeneity. I have studied the particles correlations with the typical scattering functions at different temperatures, and found out the particles motions is following the very well known regimes observed experimentally [12, 14], theoretically [4, 6, 34] and by modeling [11] for all the temperatures investigated. In a second stage, I have analyzed the particles motion and scattering function fluctuations using the standard functions, i.e. α_2 parameter and χ_4, that helped me to distinguish between the random diffusive motion typical of a liquid, from a non-Gaussian behavior typical of glassy systems. In a third stage, I have applied the standard definitions of the particles mobility [6], and, I have coupled the k-means clustering technique to infer more information on the mobility-cluster shapes and evolution in time.

I have, myself, chosen the ML technique, and I have found its results consistent with the information given by the other standard tools used in MD. Thus, adding important information on the shape of the mobile-clusters and on their evolution. I am aware that this work would need more statistical significance, but I think, given the short time available, I have been able to reach a good understanding of the problems and to contribute to test a ML method that was not previously used in such field. The learning curve has been very steep, but I have found the topic very intriguing and worth the efforts and ling time waiting after my simulations' completion. I have got help in diving on the topic and on the MD codes, but I have developed myself all the post-processing codes using the Julia programming language.

I have also used the Piz-Daint machine of the Swiss HPC-facilities, CSCS consortium, that is known for its GPUs super-computers [42]. It means, I have also gained very useful knowledge of the GPU accelerated machines and the syntax necessary to use those.

Bibliography

[1] C. Kittel, *Introduction to Solid State physics*. John Wiley Sons Inc, 2018.

[2] K. Binder and W. Kob, *Glassy materials and disordered solids: An introduction to their statistical mechanics*. World scientific, 2011.

[3] T. G. Mezger, *Applied Rheology: With Joe Flow on Rheology Road*. Austria: Anton Paar GmbH, 2015.

[4] W. Kob, *Supercooled liquids, the glass transition, and computer simulations*, 2002. DOI: 10.48550/ARXIV.COND-MAT/0212344. [Online]. Available: https://arxiv.org/abs/cond-mat/0212344.

[5] E. D. Zanotto and J. C. Mauro, 'The glassy state of matter: Its definition and ultimate fate,' *Journal of Non-Crystalline Solids*, vol. 471, pp. 490–495, 2017, ISSN: 0022-3093. DOI: https://doi.org/10.1016/j.jnoncrysol.2017.05.019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022309317302685.

[6] L. M. C. Janssen, 'Mode-coupling theory of the glass transition: A primer,' *Frontiers in Physics*, vol. 6, 2018, ISSN: 2296-424X. DOI: 10.3389/fphy.2018.00097. [Online]. Available: https://www.frontiersin.org/article/10.3389/fphy.2018.00097.

[7] J.-P. Hansen and I. R. McDonald, 'Chapter 4 - distribution function theories,' in *Theory of Simple Liquids (Fourth Edition)*, J.-P. Hansen and I. R. McDonald, Eds., Fourth Edition, Oxford: Academic Press, 2013, pp. 105–147, ISBN: 978-0-12-387032-2. DOI: https://doi.org/10.1016/B978-0-12-387032-2.00004-0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123870322000040.

[8] M. Micolaut, 'Relaxation and physical aging in network glasses: A review,' *Rep. Prog. Phys.*, vol. 79, p. 55, 2016.

[9] S. Karmakar, C. Dasgupta and S. Sastry, 'Growing length scales and their relation to timescales in glass-forming liquids,' *Annu. Rev. Condens. Matter Phys*, vol. 5, no. 1, pp. 255–284, 2014.

[10] I. Tah and S. Karmakar, 'Kinetic fragility directly correlates with the many-body static amorphous order in glass-forming liquids.,' *Physical Review Materials*, vol. 6, p. 035 601, 2022.

[11] I. Tah, A. Mutneja and S. Karmakar, 'Understanding slow and heterogeneous dynamics in model supercooled glass-forming liquids,' *ACS Omega*, vol. 6, no. 11, pp. 7229–7239, 2021, PMID: 33778237. DOI: 10.1021/acsomega.0c04831.

[12] B. Ruta, E. Pineda and Z. Evenson, 'Relaxation processes and physical aging in metallic glasses,' *J. Phys. Condens. Matter*, vol. 29, p. 503 002, 2017.

[13] T. Kirkpatrick and D. Thirumalai, 'Colloquium: Random first order transition theory concepts in biology and physics,' *Reviews of modern Phys*, vol. 87, p. 183, 2015.

[14] L. Berthier, G. Biroli, J. P. Bouchaud, L. Cipelletti, D. Masri, D. L'Hô te, F. Ladieu and M. Pierno, 'Direct experimental evidence of a growing length scale accompanying the glass transition.,' *Science*, vol. 310, p. 1797, 2005.

[15] S. Albert, T. Bauer, M. Michl, G. Biroli, J.-P. Bouchaud, A. Loidl, P. Lunkenheimer, R. Tourbot, C. Wiertel-Gasquet and F. Ladieu, 'Fifth-order susceptibility unveils growth of thermodynamic amorphous order in glass-formers.,' *Science*, vol. 352, pp. 1308–1311, 2016.

[16] J. Zausch, J. Horbach, M. Laurati, S. U. Egelhaaf, J. M. Brader, T. Voigtmann and M. Fuchs, 'From equilibrium to steady state: The transient dynamics of colloidal liquids under shear,' *Journal of Physics: Condensed Matter*, vol. 20, no. 40, p. 404 210, Sep. 2008. DOI: 10.1088/0953-8984/20/40/404210. [Online]. Available: https://doi.org/10.1088/0953-8984/20/40/404210.

[17] Z. Fan and E. Ma, 'Predicting orientation-dependent plastic susceptibility from static structure in amorphous solids via deep learning.,' *Nature Communications*, vol. 12, p. 1506, 2021.

[18] Q. Wang and A. Jain, 'A transferable machine-learning framework linking interstice distribution and plastic heterogeneity in metallic glasses.,' *Nature Communications*, vol. 10, p. 5537, 2019.

[19] A. Jaiswal, T. Egami and Y. Zhang, 'Atomic-scale dynamics of a model glass-forming metallic liquid: Dynamical crossover, dynamical decoupling, and dynamical clustering.,' *Phys. Rev. B*, vol. 91, p. 134 204, 2015.

[20] Q. Wang and L. Zhang, 'Inverse design of glass structure with deep graph neural networks.,' *Nature Communications*, vol. 12, p. 5359, 2021.

[21] J. Li, T. Chen and A. Zekiy, 'Correlative study between elastic modulus and glass formation in zrcual(x) amorphous system using a machine learning approach.,' *Applied Physics A*, vol. 127, p. 720, 2021.

[22] A. Bertheussen, *Characterization of ordered and disordered systems using molecular dynamics*, 2021. [Online]. Available: https://github.com/ambe98/Masters-Thesis.

[23] T. J. Vlugt, J. P. van der Eerden, M. Dijkstra, B. Smit and D. Frenkel, *Introduction to Molecular Simulation and Statistical Thermodynamics*. Delft, The Netherlands, 2008, ISBN: 978-90-9024432-7. [Online]. Available: http://www.phys.uu.nl/%E2%88%BCvlugt/imsst.

[24] *Clustering.jl package*, https://juliastats.org/Clustering.jl/stable/, Accessed: 2022-04-30.

[25] H. Vogel, *Phys. Z*, vol. 22, p. 645, 1921.

[26] G. S. Fulcher, *J. Amer. Ceram. Soc.*, vol. 8, p. 339, 1925.

[27] G. Tammann and W. Z. Hesse, *Anorg. Allg. Chem.*, vol. 156, p. 245, 1926.

[28] F. Stickel, E. W. Fischer and R. Richert, 'Dynamics of glass-forming liquids. i. temperature-derivative analysis of dielectric relaxation data,' *The Journal of Chemical Physics*, vol. 102, no. 15, pp. 6251–6257, 1995. DOI: 10.1063/1.469071. eprint: https://doi.org/10.1063/1.469071. [Online]. Available: https://doi.org/10.1063/1.469071.

[29] W. Gotze and L. Sjogren, 'Relaxation processes in supercooled liquids,' *Reports on Progress in Physics*, vol. 55, no. 3, pp. 241–376, Mar. 1992. DOI: 10.1088/0034-4885/55/3/001. [Online]. Available: https://doi.org/10.1088/0034-4885/55/3/001.

[30] W. Götze, *Complex dynamics of glass-forming liquids: A Mode-Coupling Theory*. Oxford, UK: Oxford University Press, 2009.

[31] L. E., 'Dynamical model of the liquid-glass transition.,' *Phys Rev A*, vol. 29, p. 2765, 1984.

[32] U. (Balucani and M. Zoppi, *Dynamics of the liquid state / Umberto Balucani and Marco Zoppi*. eng, ser. Oxford series on neutron scattering in condensed matter ; 10. Oxford : New York: Clarendon Press ; Oxford University Press, 1994, ISBN: 0198517394.

[33] F. Weysser, A. Puertas, M. Fuchs and T. Voigtmann, 'Structural relaxation of polydisperse hard spheres: Comparison of the mode-coupling theory to a langevin dynamics simulation.,' *Phys. Rev. E*, vol. 82, p. 011 504, 2010.

[34] W. Kob, M. Nauroth and F. Sciortino, 'Quantitative tests of mode-coupling theory for fragile and strong glass formers.,' *J. Non Cryst. Solids*, vol. 307-310, pp. 181–7, 2002.

[35] M. D. Ediger, 'Spatially heterogeneous dynamics in supercooled liquids,' *Annual Review of Physical Chemistry*, vol. 51, no. 1, pp. 99–128, 2000, PMID: 11031277. DOI: 10.1146/annurev.physchem.51.1.99. eprint: https://doi.org/10.1146/annurev.physchem.51.1.99. [Online]. Available: https://doi.org/10.1146/annurev.physchem.51.1.99.

[36] L. Cipelletti and E. R. Weeks, *Glassy dynamics and dynamical heterogeneity in colloids*, 2010. DOI: 10.48550/ARXIV.1009.6089. [Online]. Available: https://arxiv.org/abs/1009.6089.

[37] E. Flenner and G. Szamel, 'Dynamic heterogeneity in a glass forming fluid: Susceptibility, structure factor, and correlation length,' *Phys. Rev. Lett.*, vol. 105, p. 217 801, 21 Nov. 2010. DOI: 10.1103/PhysRevLett.105.217801. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.105.217801.

[38] K. Kim and S. Saito, 'Multiple length and time scales of dynamic heterogeneities in model glass-forming liquids: A systematic analysis of multi-point and multi-time correlations,' *The Journal of chemical physics*, vol. 138, 12A506, Mar. 2013. DOI: 10.1063/1.4769256.

[39] T. Vlugt, J. Eerden and M. Dijkstra, 'Introduction to molecular simulation and statistical thermodynamics,' Jan. 2008.

[40] S. Plimpton, 'Fast parallel algorithms for short-range molecular dynamics,' *Journal of Computational Physics*, vol. 117, no. 1, pp. 1–19, 1995, ISSN: 0021-9991. DOI: https://doi.org/10.1006/jcph.1995.1039. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002199918571039X.

[41] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott and S. J. Plimpton, 'LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales,' *Comp. Phys. Comm.*, vol. 271, p. 108 171, 2022. DOI: 10.1016/j.cpc.2021.108171.

[42] *Cscs - swiss national supercomputing centre*, https://www.cscs.ch/publications/news/2017/factsheetpizdaintoneofthemostpowerfulsupercomputersintheworld/, Accessed: 2022-06-23.

[43] T. Soddemann, B. Dünweg and K. Kremer, 'Dissipative particle dynamics: A useful thermostat for equilibrium and nonequilibrium molecular dynamics simulations,' *Phys. Rev. E*, vol. 68, p. 046 702, 4 Oct. 2003. DOI: 10.1103/PhysRevE.68.046702. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.68.046702.

[44] F. Arceri, F. P. Landes, L. Berthier and G. Biroli, *Glasses and aging: A statistical mechanics perspective*, 2020. DOI: 10.48550/ARXIV.2006.09725. [Online]. Available: https://arxiv.org/abs/2006.09725.

[45] G. Tarjus, *An overview of the theories of the glass transition*, 2010. DOI: 10.48550/ARXIV.1010.2938. [Online]. Available: https://arxiv.org/abs/1010.2938.

[46] P. Chaudhuri, L. Berthier and W. Kob, 'Universal nature of particle displacements close to glass and jamming transitions,' *Phys. Rev. Lett.*, vol. 99, p. 060 604, 6 Aug. 2007. DOI: 10.1103/PhysRevLett.99.060604. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.99.060604.

[47] D. Arthur and S. Vassilvitskii, 'How slow is the k-means method?' In *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry*, ser. SCG '06, Sedona, Arizona, USA: Association for Computing Machinery, 2006, pp. 144–153, ISBN: 1595933409. DOI: 10.1145/1137856.1137880. [Online]. Available: https://doi.org/10.1145/1137856.1137880.

[48] D. Arthur and S. Vassilvitskii, 'K-means++: The advantages of careful seeding,' in *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, New Orleans, Louisiana: Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035, ISBN: 978-0-898716-24-5.

[49] G. Malandain and J.-D. Boissonnat, 'Computing the Diameter of a Point Set,' *International Journal of Computational Geometry and Applications*, vol. 12, no. 6, pp. 489–510, 2002. [Online]. Available: https://hal.inria.fr/inria-00615026.

[50] V. Ramasubramani, B. D. Dice, E. S. Harper, M. P. Spellings, J. A. Anderson and S. C. Glotzer, 'Freud: A software suite for high throughput analysis of particle simulation data,' *Computer Physics Communications*, vol. 254, p. 107 275, 2020, ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2020.107275. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010465520300916.

[51] N. Kikuchi and J. Horbach, 'Mobile particles in an immobile environment: Molecular dynamics simulation of a binary yukawa mixture.,' *EPL*, vol. 77, p. 26 001, 2007.

[52] J. Bezanson, A. Edelman, S. Karpinski and V. B. Shah, 'Julia: A fresh approach to numerical computing,' *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017. DOI: 10.1137/141000671. [Online]. Available: https://epubs.siam.org/doi/10.1137/141000671.

[53] C. Yang, Y. Li and F. Cheng, 'Accelerating k-means on GPU with CUDA programming,' *IOP Conference Series: Materials Science and Engineering*, vol. 790, no. 1, p. 012 036, Mar. 2020. DOI: 10.1088/1757-899x/790/1/012036. [Online]. Available: https://doi.org/10.1088/1757-899x/790/1/012036.

Appendix A

Lammps Input File

The input file described in code listing A.1 is an example of how the experiment was run using the LAMMPS software to simulate a binary colloidal system described in this project. The initial.config file was either the initial configuration described in 3.2 or replaced by one of the output files from an earlier simulation in order to continue from a sample already converged into a specific temperature. Because all the temporal changes was analysed over a logarithmic timescale, the output frequency was adjusted to evenly be distributed in logarithmic time.

Code listing A.1: Lammps input file used for simulations of the binary colloidal system. Temperature was set to $T = 0.14$

```
suffix          gpu
units           lj      # Use Lennard-Jones units
atom_style      sphere  # Use Colloids
dimension       3

# Import initial configuration or previous configuration
#read_data      initial.config
read_data       sim_res/data_014_1.colloid

# Define system parameters
set             type 1 mass 1.0
set             type 2 mass 1.0

set             type 1 diameter 1.0
set             type 2 diameter 1.2

variable        t equal step

# Multi neighbor and comm for efficiency

neighbor        1 multi
neigh_modify    delay 0
comm_modify     mode multi
comm_modify     vel yes

# Colloidal potential (Yukawa) and thermostat

pair_style      hybrid/overlay yukawa 6.0 2.5 dpd/tstat 0.14 &
                0.14 1.25 34387
```

```
# The hybrid command allows to combine the yukawa potential
# and the dpd thermostat for the temperature

pair_coeff      1 1  yukawa  403.4588  3.4784
pair_coeff      1 2  yukawa 1132.0466  3.3253
pair_coeff      2 2  yukawa 3214.6338  3.1984

# dpd parameters
pair_coeff      1 1  dpd/tstat 12 1.25
pair_coeff      1 2  dpd/tstat 12 1.25
pair_coeff      2 2  dpd/tstat 12 1.25

change_box      all triclinic

fix             1 all nve

# Change output to logarithmic frequency
variable        s equal logfreq2(1,18,10)
dump            1 all custom 100 pos_014K_1.dump  id type xu yu zu x y z
dump            2 all custom 100 vel_014K_1.dump  id type vx vy vz x y z
dump_modify     1 every v_s first yes
dump_modify     2 every v_s first yes

# Define thermo-log parameters
thermo_style    multi
thermo_style    custom step time temp etotal press vol pxx pxy xy
thermo          1000

timestep        0.0083
# Restart file in case simulation stops
restart         100000 restart1.dat restart2.dat

# Run simulation
run             2000000

# Save end configuration
#write_data      data_014_1.colloid
```

Appendix B

Optimization process of static structure factor calculation

From the partial static structure factor given by equation (2.8), a system with $N_A = N_B = 8000$ will need to do calculations for $8000^2 = 64 \times 10^6$ particle pairs. For a large set of k-vectors, this can be a cumbersome calculation. The number of k-vectors, N_k, included in the set given by equation (4.1) can be approximated by the volume spanned by (n_x, n_y, n_z). Thus, with $k_{max} = 15$ and $\Delta k \approx 0.21874$, one gets

$$N_k \approx \frac{\pi}{6}\Big(\frac{k_{max}}{\Delta k}\Big)^3 \approx 170\,000. \tag{B.1}$$

The calculation heavy post-processing scripts, like calculating the static structure factor, were all written in the Julia programming language [52]. This choice was made because of Julia being a dynamic language, make it easy to write, while it at the same time runs generally fast. Additionally, Julia is developed around scientific programming, which means that there exist a wide range of open-source packages and tools to help make the post-processing efficient and fast.

Initially the static structure factor was calculated through a double for-loop, exactly as described in equation church. Because of the high number of particle pairs considered in the calculation, the algorithm ended up taking 2.1 seconds for each wave vector. In order to utilize the CPU better, the calculation needed to be vectorized. This was done by defining a matrix

$$X_{i,j} = r_{x,i} - r_{x,j}, \tag{B.2}$$

where $r_{x,i}$ is the x-position of particle i. Similarly Y and Z was defined for the two other directions. All subtractions included in the matrices was then only calculated once instead of doing it for all wave vectors. The static structure was now given as

$$S(\vec{k}) = \frac{1}{N}\Big(sum\big(\cos(k_x X + k_y Y + k_z Z)\big) + sum\big(\sin(k_x X + k_y Y + k_z Z)\big)\Big). \tag{B.3}$$

With X, Y and Z being 8000×8000 matrices, the algorithm was faster, but still far from sufficient. A nice speed-up was made through the realization that the

matrices was symmetric with zero along the diagonal. Thus, the calculation could be cut in half by only using the upper triangular part of the matrices. After converting them to arrays of 32×10^6 elements each, the calculation used 0.84 seconds per wave vector, which was around 2.5 times faster than the initial algorithm. With the limitations for single-core computing starting to appear, the next speed up was done by multiprocessing the task. Running calculations for different k-vectors on multiple threads helped to further half the estimated calculation time (for a 6-core CPU). By calculating the static structure factor for 20 wave vectors, it was estimated that each calculation now took 0.46 seconds. Thus, it would take around 22 hours to run the calculation on the total set of wave vectors. The progression is summarized in table B.1.

Table B.1: Progression on optimization of the static structure factor calculation. Total estimated calculation time is based on average time elapsed for calculations on 20 k-vectors and then scaled up to 170 000 k-vectors.

Version	Total estimated calculation time (hours)
Double for-loop	107
Vectorization	84
Upper triangular part	42
Multiprocessing	22
Utilizing the GPU	0.36 *

* Nvidia GTX 1050 Ti Max-Q

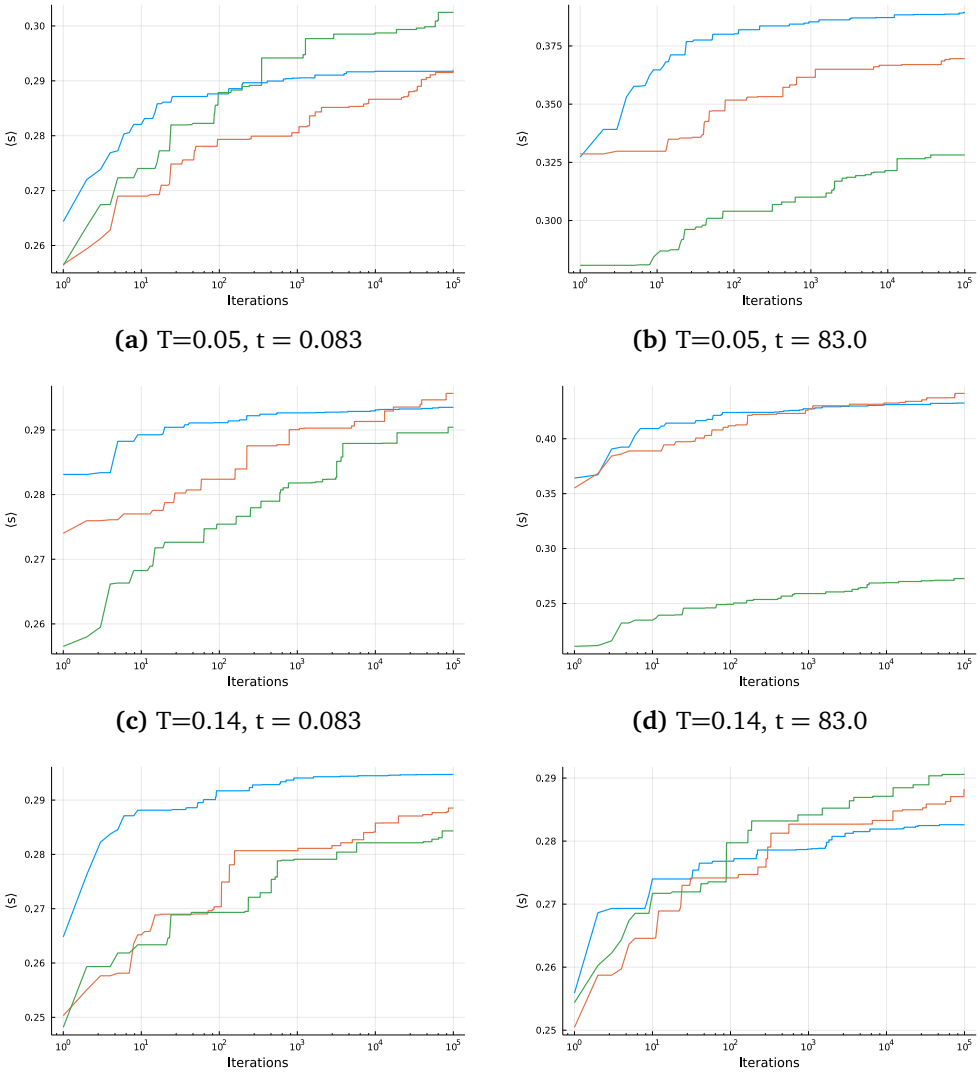
Although running the calculation for 22 hours was doable, it was not desirable to let a laptop run on 90%+ CPU utilization for that amount of time. This could have been solved by rewriting the algorithm to run on a high-performance cluster where it could have employed a bigger number of CPU-cores. Another, and potentially easier solution, was to run most of the calculations on a GPU. All the mathematical operations involved in the calculations were basic and thus allowed to run on GPU-cores. With today's GPUs packing several hundreds or thousands of cores, a huge speed-up could be gained. With use of the CUDA.jl package and a dedicated NVIDIA graphics card, the calculation time was reduced to a staggering 8 milliseconds per wave vector.

Appendix C

Convergence of k-means

To ensure accurate measurements using the k-means algorithm, a convergence analysis was performed. For a range of temperatures, times and values for κ, the best silhouette score $\langle s \rangle$ was plotted as a function of iterations with k-means on the mobile particles. The configurations were chosen to best represent the different configurations investigated in this thesis, ie. both liquid and glassy states at different timesteps. The results are shown in figure C.1, where the blue, orange and green lines represent scores for $\kappa = 10$, $\kappa = 50$ and $\kappa = 100$ respectively. Each line is given by an average over three samples.

The convergence analysis highlights one of the difficulties with using the k-means algorithm as a method to cluster the particles. The optimization problem solved by the algorithm has many local minima and thus numerous iterations are required to achieve the best possible solution. From figure C.1 we can see that many of the configurations requires up to 1×10^5 number of iterations in order to ensure a decent convergence. This is especially the case for truly homogeneous particle configurations, as the global minima is ambiguous. The system at $T = 0.14$ with $t = 83.0$ however seem to converge faster than the other configurations, expectedly because of the heterogeneous properties of the configuration. The optimization problem has a more evident global minima when the best value for κ is used on a homogeneous configuration, resulting in a fast convergence into a solution with a high silhouette score. One should note the scale of the axis, and even though some configurations increase the silhouette score by 0.5 over 1×10^5 iterations, the expected heterogeneous solution consistently holds a higher maximum silhouette score.



(a) T=0.05, t = 0.083 **(b)** T=0.05, t = 83.0

(c) T=0.14, t = 0.083 **(d)** T=0.14, t = 83.0

(e) T=1.0, t = 0.083 **(f)** T=1.0, t = 83.0

Figure C.1: Best average silhouette score for three different temperatures at two different times as a function of iterations with the k-means algorithm. The blue, orange and green lines represent scores for $\kappa = 10$, $\kappa = 50$ and $\kappa = 100$ respectively. Each line is given by an average over three samples.

Because no universal value for κ used in the k-mean algorithm could be found, the scan to find the optimal κ had to be conducted for all timesteps for all temperatures. Due to the limited timeframe of this project, the true convergence criteria were not fully met during the calculation of the silhouette scores. Because the gain of a higher number of iterations showed to be much less than the difference between a heterogeneous and a homogeneous system, and further between a good and a bad value for κ, the qualitative conclusions of the silhouette scores was believed to be correct. Argutely, a smaller number of iterations is only highlight-

ing the heterogeneous configurations, as a heterogeneous configurations seem to converge into a global minimum faster than a clusterless system. With more time and resources, a smother and more accurate calculation of the silhouette scores could have been achieved. There exist several methods for speeding up the calculations, where parallelization combined with HPC is an easy way. Alternatively, k-means has been shown to adopt a significant speed gain when run on GPU [53]. Further, an adaptive stopping criterion could have been developed to account for the variations in number of iterations required for decent convergence and thus increasing the efficiency of the method.

Appendix D

K-means clustering of the system

The Julia code described in this appendix was used to cluster the different configurations in this thesis using the k-means algorithm. The k-means algorithm from the Julia package Clustering.jl [24] was utilized and wrapped in the code written by the author. The method kmean() initialize the cluster centers according to the k-mean++ technique. The input, X_dict, is a dictionary containing position, id, type and displacements for all the mobile particles for the different timestep of a simulation done with the LAMMPS software.

Additionally the average cluster diameter is estimated after the best silhouette score is estimated. All methods incorporated in this code was described in more depth by section 3.3.2.

Code listing D.1: Code used to cluster the particles, calculate the silhouette scores and estimate the cluster diameters for each timestep in a simulation.

```
function silhouettes_scores(X_dict::Dict{Any,Any})
    #Function that calculates the silhouette scores and cluster diameter
    #for each timesteps for the mobile particles given in X_dict.
    #
    #X_dict includes particle: id, type, position, displacements

    # Array containing all the timesteps
    ts = sort(collect(keys(X_dict)))

    #Initialize arrays to store results
    scores = zeros(Float64, length(ts))     # Best silhouette score
    cluster_nums = zeros(Int64, length(ts))
    # Number of cluster that give the best score
    avg_diam = zeros(Float64, length(ts))     # Average cluster diameter

    # Loop over timesteps
    for (i,t) in enumerate(ts)
        # Extract the wrapped cordinates into a matrix
        X = Array(X_dict[t][:,[:xw,:yw,:zw]])'

        # Locate area for best k-value (Number of clusters)
        max_k = minimum([floor(Int, nrow(X_dict[t])*0.5), 150])
        N_clusters = Array(2:4:max_k)
        best_scores = zeros(Float64, length(N_clusters))
```

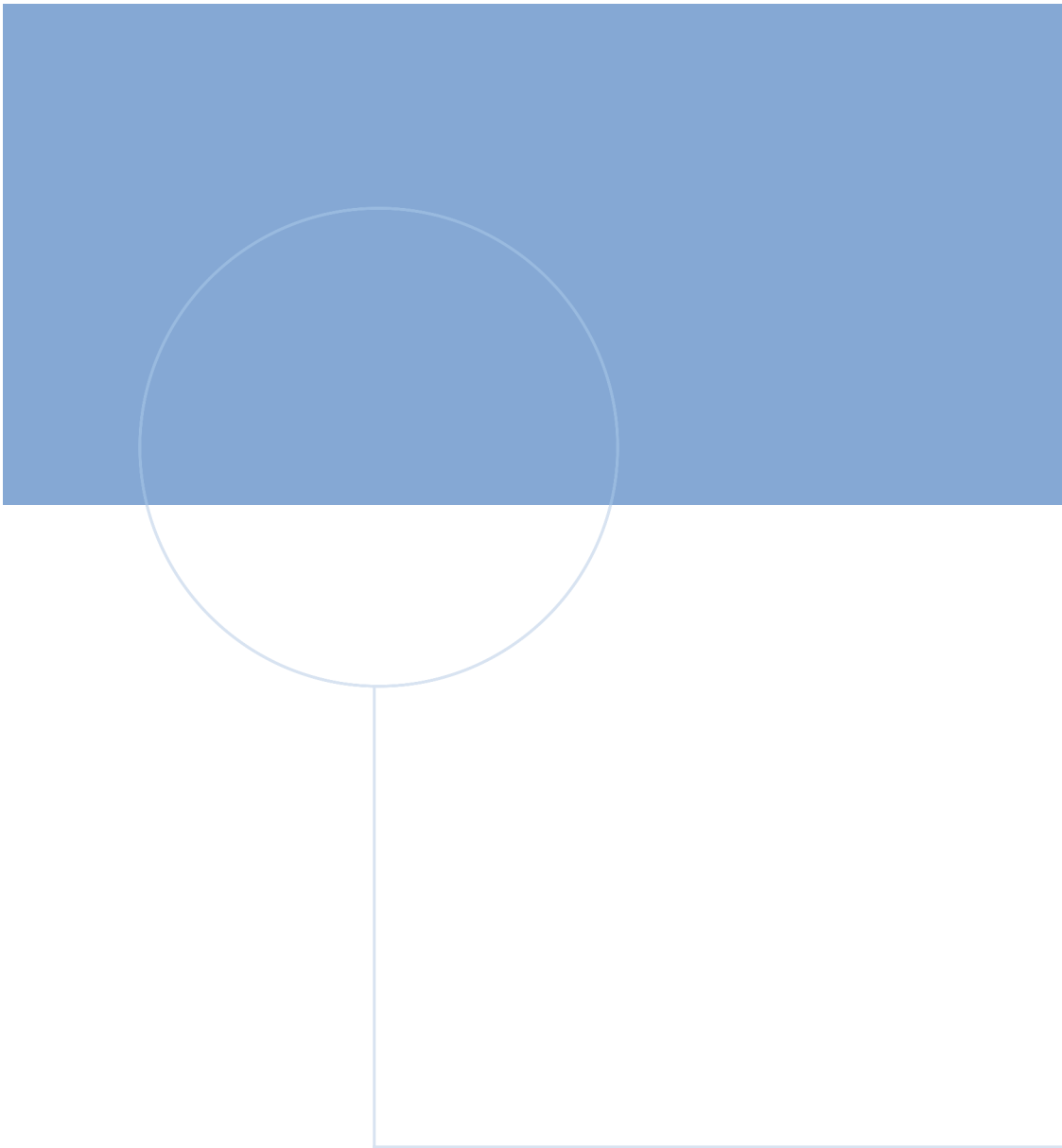
```
for (i, k) in enumerate(N_clusters)
    samples = 30
    best = zeros(1)
    #Find max silluette score for samples
    for j in 1:samples
        R = kmeans(X, k; display=:none)
        score = mean(silhouettes(R, pairwise(Euclidean(), X, dims=2)))
        if score > best[1]
            best[1] = score
            sort!(best)
        end
    end
    # For each k, store the best score of 30 samples
    best_scores[i] = mean(best)
end
# The best k should be around the k that gave the best score
k_g = N_clusters[argmax(sma(best_scores,1,1))]
best_R = kmeans(X, k_g; display=:none)

# Refine the search to find the best k-value
iterations = 500
best_score = 0
best_k = 0
lower = k_g-10>2 ? k_g-10 : 2 # Lower bound for k is 2
# Search area is the previous best k, +- 10
N_clusters = Array(lower:k_g+10)
for (i, k) in enumerate(N_clusters)
    #Find best silhouette score over 500 iterations
    for j in 1:iterations
        R = kmeans(X, k; display=:none)
        score = mean(silhouettes(R, pairwise(Euclidean(), X, dims=2)))
        if score > best_score
            best_score = score
            best_k = k
            best_R = R
        end
    end
end
# Store the best k and the respective silhouette score
scores[i] = best_score
cluster_nums[i] = best_k

# Calculate the cluster diameters for the best clustering solution, R
diameters = zeros(best_k)
# Loop over all the clusters
for k in 1:best_k
    # Filter out the particles assigned to cluster k
    Xs = X[:,findall(best_R.assignments .== k)]
    # Calculate all the pair distances and save
    # the longest intra-cluster distance
    for j in 1:(size(Xs)[2])
        diam = maximum(sqrt.((Xs[1,j:end].-Xs[1,j]).^2
        .+(Xs[2,j:end].-Xs[2,j]).^2
        .+(Xs[3,j:end].-Xs[3,j]).^2))
        if diam > diameters[k]
            diameters[k] = diam
        end
    end
end
# Store the average cluster-diameter
```

```
        avg_diam[i] = mean(diameters)

    end
    # Return the timesteps, silhouette scores, number of clusters
    # and average cluster diameter
    return ts, scores, Int.(cluster_nums), avg_diam
end
```



NTNU

Norwegian University of
Science and Technology