



**HAL**  
open science

# Improving the prediction of glassy dynamics by pinpointing the local cage

Rinske Alkemade, Frank Smallenburg, Laura Filion

► **To cite this version:**

Rinske Alkemade, Frank Smallenburg, Laura Filion. Improving the prediction of glassy dynamics by pinpointing the local cage. *The Journal of Chemical Physics*, 2023, 158 (13), pp.134512. <10.1063/5.0144822>. <hal-04309304>

**HAL Id: hal-04309304**

**<https://hal.science/hal-04309304v1>**

Submitted on 27 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Improving the prediction of glassy dynamics by pinpointing the local cage

Rinske M. Alkemade,<sup>1</sup> Frank Smallenburg,<sup>2</sup> and Laura Filion<sup>1</sup>

<sup>1</sup>*Soft Condensed Matter, Debye Institute of Nanomaterials Science, Utrecht University, Utrecht, Netherlands*

<sup>2</sup>*Université Paris-Saclay, CNRS, Laboratoire de Physique des Solides, 91405 Orsay, France*

The relationship between structure and dynamics in glassy fluids remains an intriguing open question. Recent work has shown impressive advances in our ability to predict local dynamics using structural features, most notably due to the use of advanced machine learning techniques. Here we explore whether a simple linear regression algorithm combined with intelligently chosen structural order parameters can reach the accuracy of the current, most advanced machine learning approaches for predicting dynamic propensity. To do this we introduce a method to pinpoint the cage state of the initial configuration – i.e. the configuration consisting of the average particle positions when particle rearrangement is forbidden. We find that, in comparison to both the initial state and the inherent state, the structure of the cage state is highly predictive of the long-time dynamics of the system. Moreover, by combining the cage state information with the initial state, we are able to predict dynamic propensities with unprecedentedly high accuracy over a broad regime of time scales, including the caging regime.

## I. INTRODUCTION

Understanding the relation between structure and dynamics in glassy systems has sparked extensive discussion over the last few decades<sup>1–4</sup>. While cooling or compressing a glassy system, we typically observe little to no change in structure, while at the same time observing an extreme decrease in dynamics<sup>5</sup>. This decrease in dynamics is highly heterogeneous, with increasingly large regions of slow and fast particles as a function of supercooling<sup>6–8</sup>.

One of the approaches to probe the apparent discrepancy between structure and dynamics has been the use of machine learning<sup>9–20</sup>. By capturing the local structure of particles in terms of parameters and training algorithms to predict the mobility of particles based on these parameters, the idea is that we can learn what aspects of the structure influence the heterogeneous dynamics.

Over the last two years, the quest for accurate dynamical predictions in glassy systems has led to an explosion of papers introducing new methodologies that compete in predicting the so-called dynamic propensity of simple glassy models. This propensity is defined as the average expected displacement a particle will undergo in a certain time interval when starting from a specific initial configuration<sup>21,22</sup>. This explosion started in 2020 with the work of Bapst *et al.*<sup>13</sup>, where several machine learning methods were trained to predict the dynamic propensity of a Kob-Andersen system<sup>23</sup>, with a graph neural network (GNN) performing the best. In 2021, a linear-regression-based algorithm with input parameters that captured structure over several length scales was shown to be able to rival GNNs in predicting the propensity for the same system<sup>16</sup>. Since then, several works have improved on this feat, by e.g. using physics-informed parameters as input for a deep neural network<sup>17</sup>, by modifying the loss function of a GNN to also consider relative displacements between pairs of particles<sup>18</sup>, or by designing GNNs that preserve roto-translation equivariance<sup>19</sup>. These works clearly demonstrate that careful consideration of the physics involved can aid in improving the predictive accuracy of these advanced methods. However, these neural-network based approaches still carry the downside of high complexity. In contrast, due

to the simplicity of linear methods, accurately capturing dynamics using linear regression gives a clearer perspective on what structural aspects most strongly drive glassy dynamics. This raises the question: can a clever choice of input parameters boost the performance of linear regression approaches?

Here, we show that this is indeed the case, and apply linear regression to predict the dynamic propensity of three glass-forming models: hard spheres, harmonic spheres, and the Kob-Andersen model. The main idea of our method is to consider the structure of the system during the caging regime, where each particle is confined by its neighbors in a reasonably well-defined location. By directly incorporating information about this “cage state”, we show that it is possible to drastically improve the ability of linear regression to predict dynamics, to the point where it even exceeds advanced non-linear machine learning algorithms over a wide range of time scales.

## II. MODEL, DESCRIPTORS AND PREDICTION METHOD

### A. Model

Each of our three models consists of two species of particles, labeled  $A$  and  $B$ , with different particle sizes but equal mass  $m$ . We denote the number of particles of the two species as  $N_A$  and  $N_B$ , such that the total number of particles  $N = N_A + N_B$ . Below we discuss each of the models and the statepoints at which we investigate them individually.

#### 1. Binary hard-spheres mixture

The first model we consider consists of hard-sphere particles of two diameters, denoted  $\sigma_A$  and  $\sigma_B$ . The hard-sphere potential for two particles  $i$  and  $j$  is given by

$$V^{\text{HS}}(r) = \begin{cases} \infty & \text{for } r \leq \sigma_{ij} \\ 0 & \text{for } r > \sigma_{ij}, \end{cases} \quad (1)$$

where  $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ . Here, we use a size ratio  $\sigma_B/\sigma_A = 0.85$ , and a number ratio  $N_A/N = 0.3$ . The considered packing fraction of  $\eta = 0.58$  leads to a structural relaxation time of approximately  $\tau_\alpha = 10^4\tau$ , with  $\tau$  the unit of time given by  $\tau = \sqrt{m\sigma_A^2/k_B T}$ ,  $k_B$  Boltzmann's constant and  $T$  the temperature.

## 2. Binary harmonic mixture

The binary harmonic potential is given by<sup>24,25</sup>

$$V^{\text{Har}}(r) = \begin{cases} \varepsilon \left(1 - \frac{r}{\sigma_{ij}}\right)^2 & \text{for } r \leq \sigma_{ij} \\ 0 & \text{for } r > \sigma_{ij}, \end{cases} \quad (2)$$

where again  $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ . Here, we consider the case where  $\sigma_B/\sigma_A = 1.4$ ,  $N_A/N = 0.5$ . Our state point of interest is at number density  $\rho\sigma_A^3 = 0.82$  and temperature  $k_B T/\varepsilon = 0.0045$ , where the structural relaxation time is approximately  $\tau_\alpha = 671\tau^{26}$ .

## 3. Binary Kob-Andersen mixture

The Kob-Andersen (KA) mixture consists of two particles types  $A$  and  $B$  interacting via the Lennard-Jones potential:<sup>23</sup>

$$V^{\text{KA}}(r) = 4\varepsilon_{ij} \left[ \left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^6 \right], \quad (1)$$

where  $\varepsilon_{AA} : \varepsilon_{AB} : \varepsilon_{BB} = 1 : 1.5 : 0.5$  and  $\sigma_{AA} : \sigma_{AB} : \sigma_{BB} = 1 : 0.80 : 0.88$ . Note that  $(\sigma_{AA} + \sigma_{BB})/2 \neq \sigma_{AB}$ , i.e. the system is non-additive. The composition of the system is  $N_A/N = 0.8$ . We investigate this system at number density  $\rho\sigma_A^3 = 1.203$  and temperature  $k_B T/\varepsilon_A = 0.44$ . The relaxation time of the system is approximately  $\tau_\alpha \simeq 3075\tau^{13}$ .

## B. Generating initial configurations

The HS system is simulated using event-driven molecular dynamics<sup>27,28</sup> (EDMD) in the microcanonical ensemble, i.e. at fixed number of particles  $N$ , volume  $V$  and energy  $E$ . In order to generate snapshots that can serve as initial configurations, we place  $N$  particles in the box at a reduced size, and then grow them over time until the desired packing fraction is reached<sup>29</sup>. Afterwards, the system is equilibrated for  $10\tau_\alpha$ .

To simulate both the harmonic and KA systems, we use LAMMPS<sup>30</sup>. First we initialize the system by performing a simulation in the canonical ensemble, i.e. at fixed  $N$ ,  $V$  and  $T$ , using a Nose-Hoover thermostat<sup>31</sup> at the desired temperature  $T$ . Afterwards, we equilibrate the system in the microcanonical ensemble, for  $10\tau_\alpha$ .

For each system we equilibrate 100 snapshots, where each snapshot contains 2000 particles.

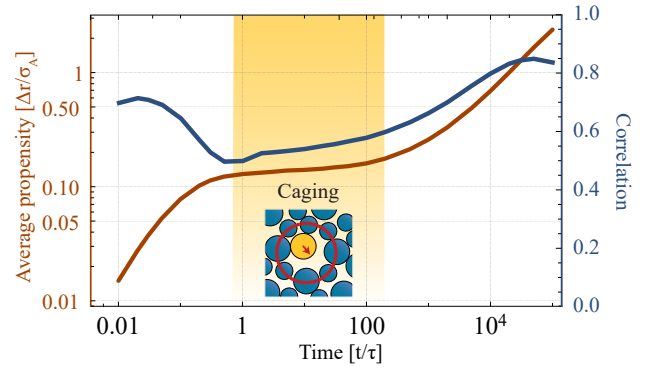


FIG. 1. Propensity averaged over all  $A$  particles in the system and correlation between the predicted and measured propensity for the same particles, plotted over time for the binary hard sphere mixture at the state point described in this paper. The propensity clearly exhibits three different dynamical regimes: In the ballistic regime, particles have not yet collided with their neighbours, and thus follow a straight ballistic trajectory. In the caging regime, particles move around for an extend period of time in the cage formed by their nearest neighbours. Finally, after particles have escaped the system enters the diffusive regime.

## C. Dynamic propensity

As a measure of the dynamical heterogeneity we use the dynamic propensity, which is a quantity that captures the average mobility of particles<sup>21,22</sup>. To measure it, we simulate the dynamical evolution of each initial configuration 50 times, using distinct initial velocities drawn from a Maxwell-Boltzmann distribution at the desired temperature, i.e. we sample the isoconfigurational ensemble<sup>21</sup>. Afterwards the dynamic propensity  $\Delta r_i(t)$  of each particle  $i$  is obtained by averaging its absolute displacement over the different runs, i.e.

$$\Delta r_i(t) = \langle |\mathbf{r}_i(t) - \mathbf{r}_i(0)| \rangle_{\text{conf}}, \quad (3)$$

where the subscript 'conf' indicates the isoconfigurational average. We measure the dynamic propensity at logarithmic time intervals between  $t/\tau = 0.01$  and  $t/\tau = 10\tau_\alpha$ . Simulations of the all the three systems are performed in a microcanonical ensemble.

## D. Structural Descriptors

To fit the dynamic propensity, we use standard ridge regression combined with structural order parameters, as previously done in Refs. 16 and 32. The structural parameters include rotationally invariant parameters that capture both the local density as well as the local  $n$ -fold symmetry. The local density is measured by using radial density functions that capture the density in a shell at distance  $r$  and with thickness  $2\delta$  from the reference particle. They are defined as

$$G_i^{(0)}(r, \delta, s) = \sum_{j \neq i, s_j = s} e^{-\frac{(r_{ij}-r)^2}{2\delta^2}}, \quad (4)$$

where  $i$  is the reference particle and  $r_{ij}$  is the absolute distance between particles  $i$  and  $j$ . The sum goes over all other particles  $j$  in the system that are part of particle species  $s$ . In this paper we include the radial functions up to the fifth minimum in the radial distribution.

The other structural descriptors we use are based on bond order parameters which express the local structure in terms of spherical harmonics<sup>33,34</sup>. To compute the parameters we first compute the complex coefficient

$$q_i^{(0)}(l, m, r, \delta) = \frac{1}{Z} \sum_{i \neq j} e^{-\frac{(r_{ij}-r)^2}{2\delta^2}} Y_l^m(\mathbf{r}_{ij}). \quad (5)$$

Here  $Y_l^m(\mathbf{r}_{ij})$  is the  $l^{\text{th}}$  spherical harmonic function and  $m$  is a function that runs from  $-l$  to  $l$ . To normalize the coefficient, we use  $Z$ , which is given by

$$Z = \sum_{i \neq j} e^{-\frac{(r_{ij}-r)^2}{2\delta^2}}. \quad (6)$$

Note that due to the exponent, just as with the radial density functions, mainly particles that are within the shell  $r - \delta$  to  $r + \delta$  contribute to  $q_i^{(0)}(l, m, r, \delta)$ . The parameters are made rotationally invariant by summing over all possible value of  $m$ :

$$q_i^{(0)}(l, r, \delta) = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{m=l} |q_i^{(0)}(l, m, r, \delta)|^2}. \quad (7)$$

Note that the  $q_i^{(0)}(l, r, \delta)$  will mainly pick-up the  $l$ -fold symmetry of the particles structure in each shell.

In Ref. 16 it was shown that the prediction of the dynamic propensity via linear regression can be improved when not only the structural parameters of the reference particle itself are included, but also structural parameters averaged over nearest neighbours. These averaged structural parameters are obtained via the following recursive formula:

$$x_i^{(n)} = \frac{\sum_{j:r_{ij}<r_c} x_j^{(n-1)} e^{-r_{ij}/r_c}}{\sum_{j:r_{ij}<r_c} e^{-r_{ij}/r_c}}. \quad (8)$$

Here  $x_i^{(n)}$  are the structural parameters (which can be both radial density or bond order parameters) of order  $n$  for particle  $i$ . The sum goes over all particles within a certain radius  $r_c$  as seen from the reference particle  $i$ . This  $r_c$  is chosen to be located at the second minimum of the radial distribution function, although it was shown in Ref. 16 that its exact value has no substantial influence on the order parameters. Boattini *et al* showed including parameters up to three generations significantly improves the predictions. Further information on the descriptors can be found in the Supplementary Information (SI).

### III. RESULTS

A number of recent studies have made significant progress in predicting the dynamic propensity of particles in glassy fluids based on local structural information using a variety of

machine learning algorithms<sup>13,16-19</sup>. The accuracy of such predictions is typically evaluated using the Pearson correlation coefficient<sup>35</sup> between the predicted and measured dynamic propensities. A typical example of this correlation as a function of time is shown in Fig. 1, where we used linear regression in combination with structural order parameters to predict the propensity in the binary hard-sphere mixture we consider here<sup>32</sup>. Interestingly, the correlation is weak in three distinct regimes. Two of these regimes are trivial. The first corresponds to the very short-time regime, where dynamics are dominated by the random choice of initial velocities. Especially for hard spheres, the motion of the particles in this regime is entirely unaffected by local structure, and hence fully unpredictable. The second trivial regime that is hard to predict occurs at long time scales  $t \gg \tau_\alpha$ , where the system loses memory of its initial configuration. The intriguing regime, where the correlations are weak, lies at intermediate time scales, and corresponds to the times where particles are trapped inside their local cages, as also indicated by the behavior of the average propensity in Fig. 1. This lower performance of machine-learned algorithms for predicting propensity in the caging regime is not unique to the results of Fig. 1, but has been observed in a variety of studies involving different machine learning methods and different ways to describe the system<sup>13,16,17,32</sup>, although recent, more advanced machine learning methods have improved significantly the correlation in this regime<sup>18,19</sup>.

Observing the weak correlations in the caging regime, a natural question to ask is then: what structural information do we need to include in order to make accurate predictions about the dynamics of our system in this regime? Since during the caging regime, particles on average move around the so called 'cage center', a rather obvious choice would be to consider the expected average distance between a particle's initial position  $\mathbf{r}_i^{\text{init}}$  and its typical position while in its cage  $\mathbf{r}_i^{\text{cage}}$ . Since the dynamic propensity measures the averaged absolute distance travelled by particles, we would expect this distance between the initial and mean cage position  $\Delta r_i^{\text{cage}} = |\mathbf{r}_i^{\text{cage}} - \mathbf{r}_i^{\text{init}}|$  to be a good predictor of the dynamic propensity. While this statement seems logically trivial, an important question is: what is a good estimate of the cage center of a given particle?

A common approach in glass literature for predicting dynamical behavior in a glassy system is to quench the system to its inherent state (IS) (see e.g.<sup>12,17,36,37</sup>). The inherent state of a configuration is defined as the local potential energy minimum that one obtains via a rapid energy minimization<sup>38</sup>. Note, however, that this is not exactly the same as the average position of a particle in its cage, as the average should include the effects of thermal fluctuations as well. As such, a natural alternative choice to estimate the position of the cage center is to simply take the average position of each particle, under the constraint that no significant particle rearrangements have happened. We will refer to this second option as the cage state (CS). To explore these two definitions of  $\Delta r_i^{\text{cage}}$ , we calculate for each of our initial configurations both the inherent state and cage state positions.

For the inherent state, we use the FIRE algorithm proposed by Bitzek *et al.*<sup>39</sup> to minimize the potential energy. In the

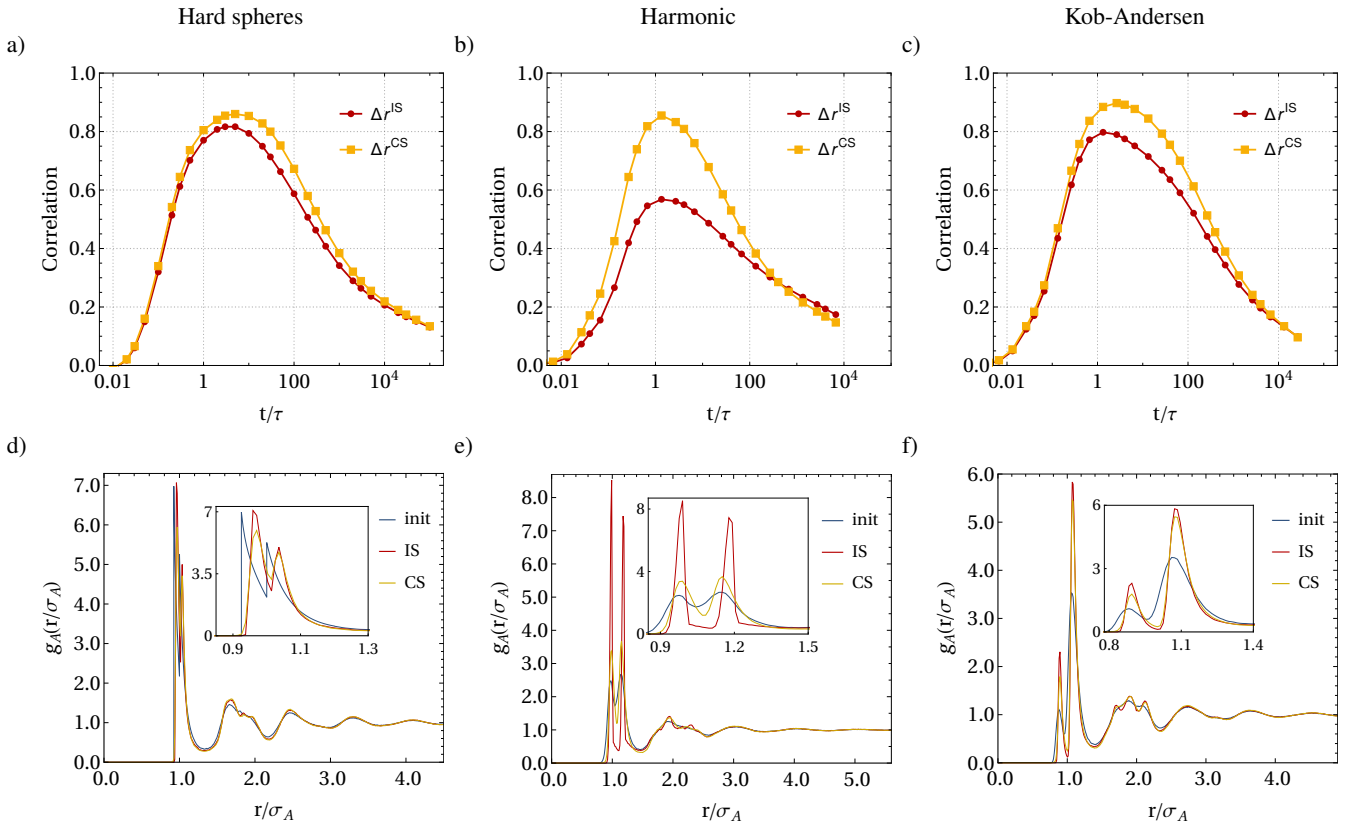


FIG. 2. **a,b,c**) The Pearson coefficient plotted over time between the propensity and absolute difference between the positions of particles in the initial configuration and the inherent state/cage state ( $\Delta r^{\text{IS/CS}}$ ) for the A-particles in hard spheres (a), harmonic spheres (b), and Kob-Andersen (c). **d,e,f**) The total radial distribution function around A-type particles in the initial configuration, the inherent state, and the cage state, for the same three models.

case of hard spheres, we apply an effective logarithmic interaction potential proposed by Arceri *et al.*<sup>40</sup> in order to obtain an effective inherent state. Specifically, we use the effective interaction potential  $V^{\text{eff}}(r) = -k_B T \log(r - \sigma_{ij})$ , with both the forces and interaction energy truncated and shifted to zero at a cutoff radius  $r_c = 1.35\sigma_j$ .

To obtain the cage state, we use a Monte Carlo (MC) simulation in the canonical ensemble that measures the average positions of each particle  $\mathbf{r}_i^{\text{CS}} = \langle \mathbf{r}_i^c \rangle$ , while restricting the movement of each particle to ensure that it stays inside its initial cage. In order to do this, we reject all MC moves that would move the center of a particle outside of its original Voronoi cell. Since we consider binary systems, we use an approximate definition of a Voronoi cell which takes into account the particle sizes. In particular, the approximate Voronoi cell for particle  $i$  is defined as the collection of points  $\mathbf{R}$  for which

$$\frac{|\mathbf{R} - \mathbf{r}_i^{\text{init}}|}{\sigma_i} < \frac{|\mathbf{R} - \mathbf{r}_j^{\text{init}}|}{\sigma_j} \quad \forall j \in \mathcal{N}(i), \quad (9)$$

where  $\mathbf{r}_i^{\text{init}}$  is the position of particle  $i$  in the initial configuration and  $\mathcal{N}(i)$  are the nearest neighbours of particle  $i$  determined by the SANN algorithm<sup>41</sup>. For the KA mixture, we set  $\sigma_i = \sigma_{ii}$ , since the individual particle sizes are ill-defined. Note that since the restrictions on the particle positions elim-

inate the possibility of long-time diffusion, short simulations are sufficient to sample the restricted phase space. Here, we use MC simulations of  $5 \cdot 10^5$  initial steps and  $10^6$  measuring steps.

Note that as an alternative to confining each particle to its initial Voronoi cell, we have also explored the possibility of instead confining each particle to a spherical region with a fixed radius  $r_c$ . When the size of this sphere is chosen to be close to the size of the particle (and hence similar to the size of the Voronoi cell), we find essentially the same results (as shown in the SI).

These two approaches give us two results for  $\Delta r^{\text{cage}}$ , which we denote  $\Delta r^{\text{IS(CS)}}$  for the inherent (cage) state. To compare how effective these measurements are at capturing the actual cage center, we plot in Fig. 2abc the Pearson correlation between the dynamic propensity and both definitions of  $\Delta r^{\text{cage}}$ , for all three glass model systems. Note that since here we consider only a single parameter at a time, no linear regression is required to obtain a correlation. As expected, both the inherent state and the cage state provide significant information about the expected position of a particle during the caging regime. However, clearly  $\Delta r^{\text{CS}}$  is a better predictor of the dynamic propensity in all three cases, reaching correlations stronger than 0.8. This is significantly higher than the results shown in Fig. 1, despite being based only on a single

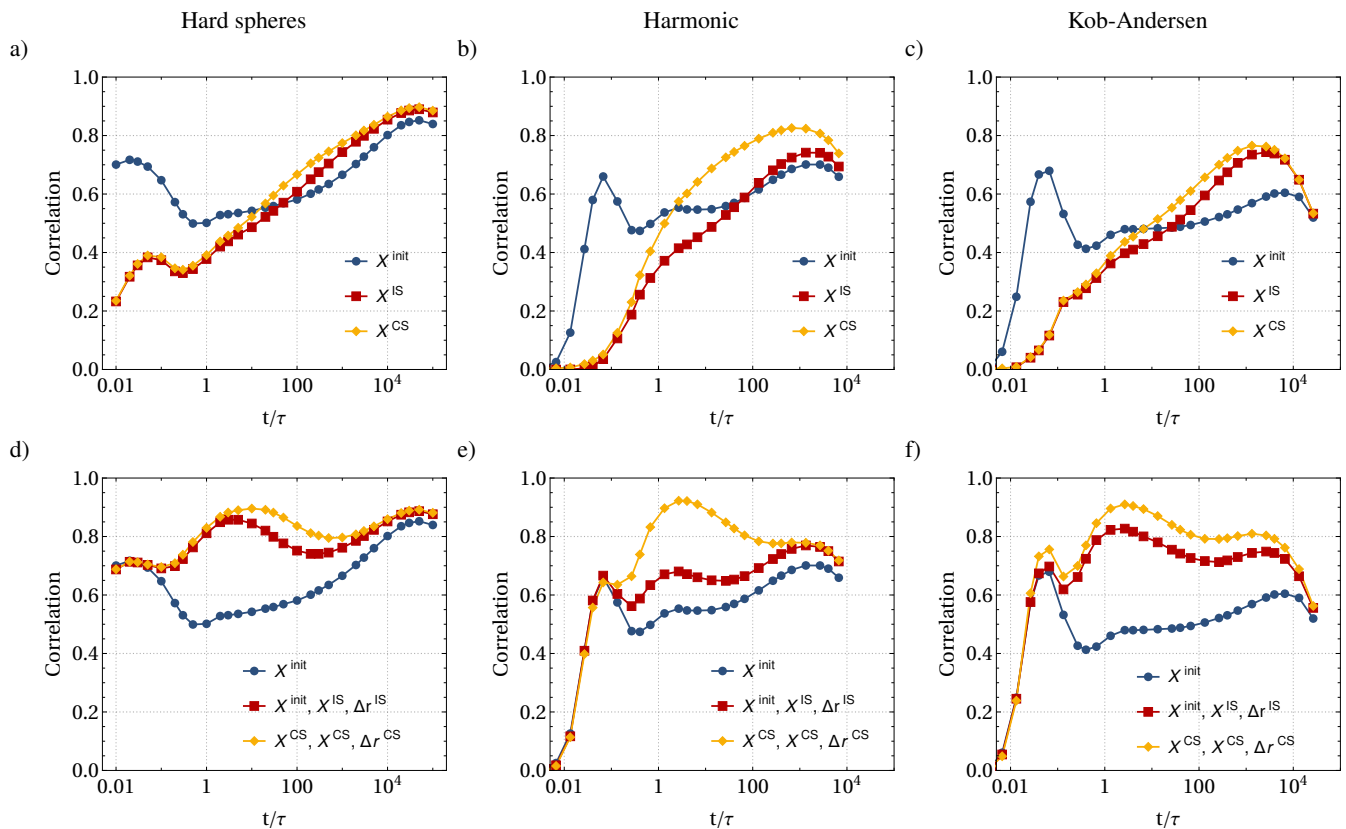


FIG. 3. **a,b,c**) Pearson correlation coefficient between the dynamic propensity and the prediction of a linear regression model trained on the structural parameters evaluated for the initial-, the inherent- and the cage-state coordinates (i.e.  $\mathbf{X}^{\text{init}}$ ,  $\mathbf{X}^{\text{IS}}$ ,  $\mathbf{X}^{\text{CS}}$ ), for the A-particles in hard spheres (a), harmonic spheres (b), and Kob-Andersen (c). **d,e,f**) Correlation between the measured propensity and the predicted propensity by a linear regression model trained on either only the structural parameters based on the initial positions, or the set of input parameters given by  $\{\mathbf{X}^{\text{init}}, \mathbf{X}^{\text{IS/CS}}, \Delta r^{\text{IS/CS}}\}$ .

variable. Evidently, the simple averaging method we used to obtain the cage state indeed manages to eliminate the thermal fluctuations that are present in the initial configuration, without losing significant information about the underlying cage structure. The inherent state performs less well, in particular in the case of harmonic spheres. To shed light on the reason behind this, we measure the radial distribution function  $g(r)$  of the initial, inherent, and cage state configurations and compare them in Fig. 2def. In the harmonic and KA models, both the inherent state and cage state increase the degree of local structure in the system, resulting in higher peaks in  $g(r)$  – and this degree of additional ordering is stronger for the inherent state than for the cage state. This implies that the inherent state quench pushes the system significantly further away from its local structure than the system would normally sample. In contrast, the cage state procedure only takes into account configurations that the system samples during (constrained) thermal fluctuations. Hence, it is perhaps not surprising that the cage state better reflects the expected dynamics of our systems.

We now examine whether knowledge of the cage structure can help us make more accurate predictions of the dynamic propensity even outside the caging regime. To this end, we start with the structural order parameters and linear regression

approach described in Refs. 16 and 32. In particular, for each particle, we define a set of approximately 1000 structural parameters describing their local structural environment, and use standard ridge regression in order to fit the dynamic propensity as a function of these structural parameters. A full description of these structural order parameters for each model is provided in the SI. Note that this is the same approach as we used for the data in Fig. 1. As a basis for calculating the local structural descriptors, we now use either the initial, inherent, or cage state, resulting in three sets of structural descriptors for each particle:  $\mathbf{X}^{\text{init}}$ ,  $\mathbf{X}^{\text{IS}}$ , and  $\mathbf{X}^{\text{CS}}$ . We then examine how these different sets of input data influence our ability to predict dynamic propensity. The results are shown in Fig. 3abc for the three model systems. In all cases, at short times the initial state of the system provides the best input for predicting propensities. This is understandable, since only the initial state contains information about the exact particle environments in the limit  $t \rightarrow 0$ . At longer time scales, both the inherent and cage state structures outperform the initial state, with the cage state always outperforming the inherent state. This further supports the observation that the cage state, as determined by our MC approach, represents an excellent approximation for the underlying structure of our system when it comes to understanding its dynamics on time scales around

the structural relaxation time.

The obvious next step is to combine this information with knowledge of  $\Delta r^{\text{cage}}$ , which we already know provides a strong prediction of dynamics in the caging regime. In Fig. 3def, we plot the correlation between the dynamic propensity and predictions based on linear regression, combining as input  $\mathbf{X}^{\text{init}}$ ,  $\mathbf{X}^{\text{cage}}$ , and  $\Delta r^{\text{cage}}$ , for both cage definitions. Overall, we observe a massive improvement in our ability to predict propensity at all time scales beyond the ballistic regime.

The predictive power of our models using information about the cage state is particularly impressive compared to past results. Specifically, the state point of the KA mixture we study here has been used in a variety of previous studies where new methodologies were introduced to attempt accurate prediction of the dynamic propensity. In Fig. 4, we plot our results from Fig. 3f and compare them to the results of Refs. 13, 16–19. Note that the comparison between different predictions is complicated by significant differences in the datasets we used here and the dataset from the work of Bapst *et al.*<sup>13</sup>, which was also used by Boattini, Jung, and Pezziccolo and their respective co-workers<sup>16,17,19</sup>. In particular, for our dataset, we determine dynamics propensities for a set of initial snapshots that are all taken at the same density, and we always measure the dynamic propensity at a fixed set of times. In contrast, the authors of Ref. 13 took the initial snapshots from constant-pressure simulations, and hence the configurations vary slightly in density. Additionally, for each initial configuration, they measured the dynamic propensity at different time intervals, based on the decay of the intermediate scattering function for that specific snapshot. In Fig. 4b, we illustrate this difference by plotting for each of the initial configurations in the training set of Ref. 13 the average propensity as a function of time, and color each point based on the corresponding system density. Hence, in our comparison in Fig. 4a, the time for each point based on this dataset is an average time. Clearly, there is a significant correlation between the density of the snapshot and the mean dynamic propensity, and a significant spread in acquisition times. Additionally, the spread in mean propensities is significantly lower than in our own data (shown in black in Fig. 4b), due to the grouping of initial configurations based on their structural relaxation instead of time. It should be noted that this grouping is not fully consistent with Eq. 3, and implies that some dynamical information is already included in the input data.

Evidently, including information about the cage state – captured in our parameters  $\mathbf{X}^{\text{CS}}$  and  $\Delta r^{\text{CS}}$  – allows our linear regression approach to outperform the current state-of-the-art machine learning methods for predicting dynamic propensity over a wide range of time scales. The GNN-based methods do outperform our predictions in the very short-time regime, where the dynamics are likely dominated by the forces that act on the particles in the initial configuration. This may be an indication that these instantaneous forces cannot be directly recovered from our set of input parameters. In contrast, the three GNN-based approaches plotted (in blue) in Fig. 4 all have information about the relative positions of particles with respect to their neighbors included in the input graph, and hence might be able to learn the net forces on the parti-

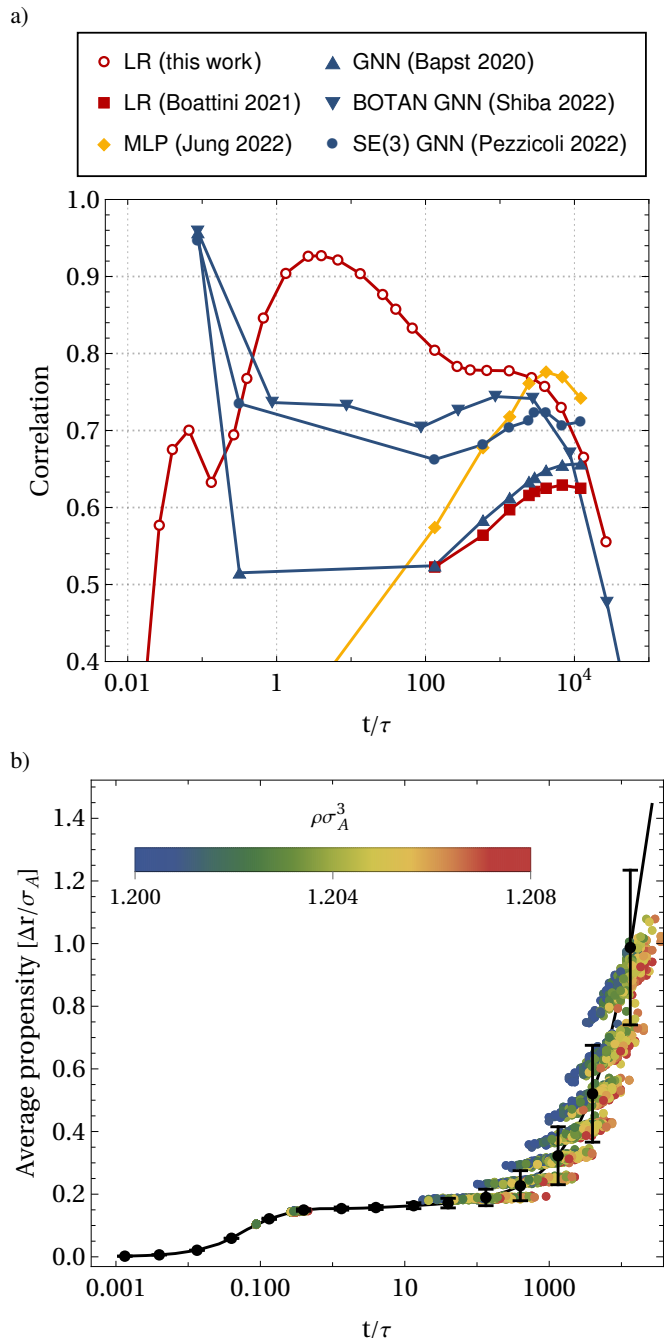


FIG. 4. a) Pearson correlation coefficient between the dynamic propensity and the prediction for A particles in a KA system at the statepoint used in this paper, made by different models: linear regression (LR) by Boattini *et al.*<sup>16</sup>, multi-layer perceptron (MLP) by Jung *et al.*<sup>17</sup>, graph neural network (GNN) by Bapst *et al.*<sup>13</sup>, BOnd TArgetting Network (BOTAN) GNN by Shiba *et al.*<sup>18</sup> and SE(3)-equivariant GNN by Pezzicoli *et al.*<sup>19</sup>. b) Average dynamic propensity per snapshot in the dataset of Ref. 13, which was also used in Refs. 16, 17, and 19. For each initial snapshot in the training set of Ref. 13, we plot the mean propensity in that snapshot at the times reported for that snapshot, and color the point based on the system density. For comparison, we also plot in black the mean propensity taken from our own dataset, where the times and density are the same for each initial configuration. The error bars indicate  $\pm 2$  times the standard deviation of the mean propensity at each time.

cles with high accuracy. At very long time scales, where the motion of particles becomes more diffusive, the results of the different machine learning approaches appear to converge to similar performance – with the physics-informed approach by Jung *et al.*<sup>17</sup> performing the strongest in this regime.

To test whether the improved performance of the GNNs at short time scales is indeed due to them directly learning the initial forces, we adapted our method to also include the net force on each particle in the initial configuration. In Fig. 5 we show the correlation between the dynamic propensity and the initial force in blue, and indeed we see a strong peak at short times. Adding this information into our input data for the full linear regression model, we indeed see a significant improvement at short times (black line). However, the GNN-based methods still achieve a significantly higher accuracy in this regime, and hence appear to be capable of learning more about the short time dynamics than simply the instantaneous forces.

In Fig. 5 we also break down our prediction of the dynamic propensity into the different relevant aspects of the input data. Overall, we see that the dynamic propensity at short times is indeed dominated by the forces, at intermediate times by the initial distance to the cage center, and at long times by the structural features of the cage state. While this is in principle not surprising, it is impressive to see that the vast majority of the variation in propensity at short and intermediate time intervals can be explained by just two simple measurements. Moreover, the predictive ability of the cage state structure at long times is impressive as well, and demonstrates that using the cage state, rather than the initial state to make long-time predictions is an excellent strategy. Note that outside of short times, adding in information on the initial state on top of the cage state is fully irrelevant.

#### IV. CONCLUSIONS

In short, we have demonstrated that the behavior of the dynamic propensity of glassy fluids can be predicted to high accuracy by using information about the cage state of the initial configuration. This cage state is defined as the set of coordinates describing the average position of the particles when the system is constrained to ensure particles cannot escape their local cages. Combining this information with a simple linear-regression-based algorithm, we can predict dynamic propensities with accuracies that rival or exceed current state-of-the-art machine learning methods at nearly all times. This suggests that the cage state could be a helpful tool for further studying the underlying structure of glassy fluids.

#### REFERENCES

#### ACKNOWLEDGEMENTS

The authors would like to thank Marjolein de Jager for many discussions. L.F. acknowledges funding from NWO for a Vidi grant (Grant No. VI.VIDI.192.102)

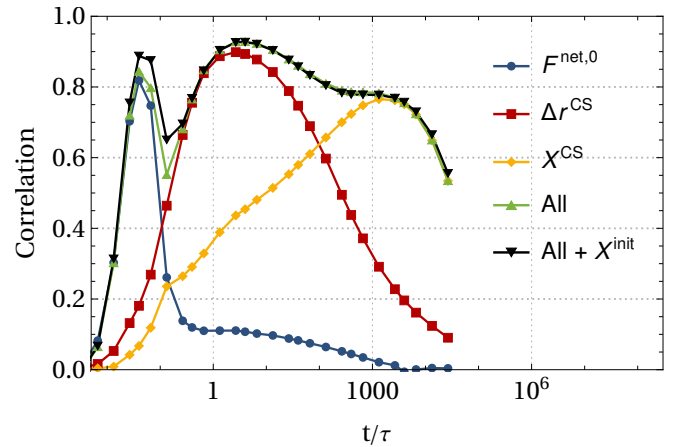


FIG. 5. Pearson correlation coefficient between measured and predicted propensity over time, shown for linear regression models trained on different subsets of structural parameters for  $A$  particles in a KA system.  $F^{\text{net},0}$  is the absolute net force that a particle feels at  $t = 0$ ,  $\Delta r^{\text{CS}}$  is the absolute difference between the positions of particles in the initial configuration and the cage state and  $\mathbf{X}^{\text{CS}}$  are the structural parameters evaluated for the cage state. Note that the green line (“All”) is the linear regression model trained on the combination of  $\{F^{\text{net},0}, \Delta r^{\text{CS}}, \mathbf{X}^{\text{CS}}\}$ , while the black line (“All +  $\mathbf{X}^{\text{init}}$ ”) also includes the structural parameters of the initial configuration in the structural dataset.

#### DATA AVAILABILITY STATEMENT

Raw data associated with the measurements of the dynamical propensity, as well as all figure data, will be published at the following url [??] prior to publication of the paper.

- <sup>1</sup>A. Cavagna, “Supercooled liquids for pedestrians,” *Phys. Rep.* **476**, 51–124 (2009).
- <sup>2</sup>L. Berthier and G. Biroli, “Theoretical perspective on the glass transition and amorphous materials,” *Rev. Mod. Phys.* **83**, 587 (2011).
- <sup>3</sup>C. P. Royall and S. R. Williams, “The role of local structure in dynamical arrest,” *Phys. Rep.* **560**, 1–75 (2015).
- <sup>4</sup>H. Tanaka, H. Tong, R. Shi, and J. Russo, “Revealing key structural features hidden in liquids and glasses,” *Nat. Rev. Phys.* **1**, 333–348 (2019).
- <sup>5</sup>M. D. Ediger, C. A. Angell, and S. R. Nagel, “Supercooled liquids and glasses,” *J. Phys. Chem.* **100**, 13200–13212 (1996).
- <sup>6</sup>M. Ediger, “Can density or entropy fluctuations explain enhanced translational diffusion in glass-forming liquids?” *J. Non-Cryst. Solids* **235**, 10–18 (1998).
- <sup>7</sup>L. Berthier and G. Biroli, “Glasses and aging, a statistical mechanics perspective on,” in *Encyclopedia of Complexity and Systems Science* (2020).
- <sup>8</sup>R. Candelier, A. Widmer-Cooper, J. K. Kummerfeld, O. Dauchot, G. Biroli, P. Harrowell, and D. R. Reichman, “Spatiotemporal hierarchy of relaxation events, dynamical heterogeneities, and structural reorganization in a supercooled liquid,” *Phys. Rev. Lett.* **105**, 135702 (2010).
- <sup>9</sup>Z.-Y. Yang, D. Wei, A. Zaccone, and Y.-J. Wang, “Machine-learning integrated glassy defect from an intricate configurational-thermodynamic space,” *Phys. Rev. B* **104**, 064108 (2021).
- <sup>10</sup>S. Ciarella, M. Chiappini, E. Boattini, M. Dijkstra, and L. Janssen, “Dynamics of supercooled liquids from static averaged quantities using machine learning,” *arXiv preprint arXiv:2212.09338* (2022).
- <sup>11</sup>E. D. Cubuk, S. S. Schoenholz, J. M. Rieser, B. D. Malone, J. Rottler, D. J. Durian, E. Kaxiras, and A. J. Liu, “Identifying structural flow defects in disordered solids using machine-learning methods,” *Phys. Rev. Lett.* **114**, 108001 (2015).

- <sup>12</sup>S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, “A structural approach to relaxation in glassy liquids,” *Nat. Phys.* **12**, 469–471 (2016).
- <sup>13</sup>V. Bapst, T. Keck, A. Grabska-Barwińska, C. Donner, E. D. Cubuk, S. S. Schoenholz, A. Obika, A. W. Nelson, T. Back, D. Hassabis, *et al.*, “Unveiling the predictive power of static structure in glassy systems,” *Nat. Phys.* **16**, 448–454 (2020).
- <sup>14</sup>E. Boattini, S. Marín-Aguilar, S. Mitra, G. Foffi, F. Smallenburg, and L. Fillion, “Autonomously revealing hidden local structures in supercooled liquids,” *Nat. Commun.* **11**, 1–9 (2020).
- <sup>15</sup>J. Paret, R. L. Jack, and D. Coslovich, “Assessing the structural heterogeneity of supercooled liquids through community inference,” *J. Chem. Phys.* **152**, 144502 (2020).
- <sup>16</sup>E. Boattini, F. Smallenburg, and L. Fillion, “Averaging local structure to predict the dynamic propensity in supercooled liquids,” *Phys. Rev. Lett.* **127**, 088007 (2021).
- <sup>17</sup>G. Jung, G. Biroli, and L. Berthier, “Predicting dynamic heterogeneity in glass-forming liquids by physics-informed machine learning,” arXiv preprint arXiv:2210.16623 (2022).
- <sup>18</sup>H. Shiba, M. Hanai, T. Suzumura, and T. Shimokawabe, “Unraveling intricate processes of glassy dynamics from static structure by machine learning relative motion,” arXiv preprint arXiv:2206.14024 (2022).
- <sup>19</sup>F. S. Pezzicoli, G. Charpiat, and F. P. Landes, “SE (3)-equivariant graph neural networks for learning glassy liquids representations,” arXiv preprint arXiv:2211.03226 (2022).
- <sup>20</sup>D. Coslovich, R. L. Jack, and J. Paret, “Dimensionality reduction of local structure in glassy binary mixtures,” *J. Chem. Phys.* **157**, 204503 (2022).
- <sup>21</sup>A. Widmer-Cooper, P. Harrowell, and H. Fynewever, “How reproducible are dynamic heterogeneities in a supercooled liquid?” *Phys. Rev. Lett.* **93**, 135701 (2004).
- <sup>22</sup>A. Widmer-Cooper and P. Harrowell, “On the study of collective dynamics in supercooled liquids through the statistics of the isoconfigurational ensemble,” *J. Chem. Phys.* **126**, 154503 (2007).
- <sup>23</sup>W. Kob and H. C. Andersen, “Testing mode-coupling theory for a supercooled binary Lennard-Jones mixture I: The van Hove correlation function,” *Phys. Rev. E* **51**, 4626 (1995).
- <sup>24</sup>D. J. Durian, “Foam mechanics at the bubble scale,” *Phys. Rev. Lett.* **75**, 4780 (1995).
- <sup>25</sup>L. Berthier and T. A. Witten, “Compressing nearly hard sphere fluids increases glass fragility,” *Euro. Phys. Lett.* **86**, 10001 (2009).
- <sup>26</sup>I. Tah, S. A. Ridout, and A. J. Liu, “Fragility in glassy liquids: A structural approach based on machine learning,” *J. Chem. Phys.* **157**, 124501 (2022).
- <sup>27</sup>D. C. Rapaport, “The Event-Driven Approach to N-Body Simulation,” *Prog. Theor. Phys. Supp.* **178**, 5–14 (2009).
- <sup>28</sup>F. Smallenburg, “Efficient event-driven simulations of hard spheres,” *Euro. Phys. J. E* **45**, 22 (2022).
- <sup>29</sup>A. Donev, S. Torquato, and F. H. Stillinger, “Neighbor list collision-driven molecular dynamics simulation for nonspherical hard particles. I. algorithmic details,” *J. Comput. Phys.* **202**, 737–764 (2005).
- <sup>30</sup>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.* **271**, 108171 (2022).
- <sup>31</sup>D. J. Evans and B. L. Holian, “The Nose–Hoover thermostat,” *J. Chem. Phys.* **83**, 4069–4074 (1985).
- <sup>32</sup>R. M. Alkemade, E. Boattini, L. Fillion, and F. Smallenburg, “Comparing machine learning techniques for predicting glassy dynamics,” *J. Chem. Phys.* **156**, 204503 (2022).
- <sup>33</sup>P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, “Bond-orientational order in liquids and glasses,” *Phys. Rev. B* **28**, 784 (1983).
- <sup>34</sup>W. Lechner and C. Dellago, “Accurate determination of crystal structures based on averaged local bond order parameters,” *J. Chem. Phys.* **129**, 114707 (2008).
- <sup>35</sup>The Pearson correlation coefficient is a measure of the degree of linear correlation between two variables. If the Pearson correlation  $\text{cor}(\mathbf{X}, \mathbf{Y})$  between two datasets  $\mathbf{X}$  and  $\mathbf{Y}$  is equal to 1, the two variables are perfectly related via a linear function with positive slope. A value of 0 indicates no linear correlation, while a value of -1 would indicate perfect negative linear correlation.
- <sup>36</sup>A. Widmer-Cooper and P. Harrowell, “Predicting the long-time dynamic heterogeneity in a supercooled liquid on the basis of short-time heterogeneities,” *Phys. Rev. Lett.* **96**, 185701 (2006).
- <sup>37</sup>H. Tong and H. Tanaka, “Revealing hidden structural order controlling both fast and slow glassy dynamics in supercooled liquids,” *Phys. Rev. X* **8**, 011041 (2018).
- <sup>38</sup>F. H. Stillinger and T. A. Weber, “Hidden structure in liquids,” *Phys. Rev. A* **25**, 978 (1982).
- <sup>39</sup>E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, “Structural relaxation made simple,” *Phys. Rev. Lett.* **97**, 170201 (2006).
- <sup>40</sup>F. Arceri and E. I. Corwin, “Vibrational properties of hard and soft spheres are unified at jamming,” *Phys. Rev. Lett.* **124**, 238002 (2020).
- <sup>41</sup>J. A. van Meel, L. Fillion, C. Valeriani, and D. Frenkel, “A parameter-free, solid-angle based, nearest-neighbor algorithm,” *J. Chem. Phys.* **136**, 234107 (2012).