PAPER • OPEN ACCESS

Dynamics of supercooled liquids from static averaged quantities using machine learning

To cite this article: Simone Ciarella *et al* 2023 *Mach. Learn.: Sci. Technol.* **4** 025010

View the article online for updates and enhancements.

MACHINE LEARNING
Science and Technology

PAPER

Dynamics of supercooled liquids from static averaged quantities using machine learning

Simone Ciarella[1,2,]* ⓘ , Massimiliano Chiappini[3], Emanuele Boattini[3], Marjolein Dijkstra[3] ⓘ and Liesbeth M C Janssen[2,]* ⓘ

[1] Laboratoire de Physique de l'Ecole Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France
[2] Soft Matter and Biological Physics, Department of Applied Physics, Eindhoven University of Technology, Den Dolech 2, 5600 MB Eindhoven, The Netherlands
[3] Soft Condensed Matter, Debye Institute for Nanomaterials Science, Princetonplein 1, Utrecht 3584 CC, The Netherlands
* Authors to whom any correspondence should be addressed.

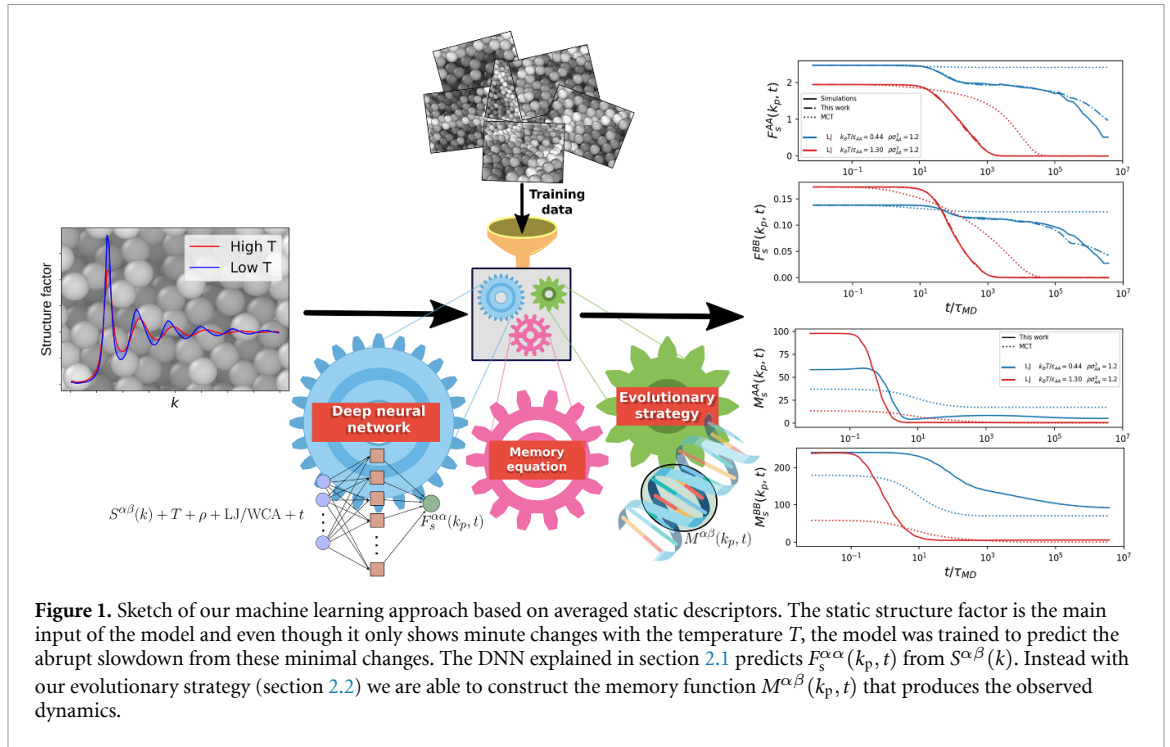E-mail: simoneciarella@gmail.com and L.M.C.Janssen@tue.nl

Abstract

We introduce a machine-learning approach to predict the complex non-Markovian dynamics of supercooled liquids from static averaged quantities. Compared to techniques based on particle propensity, our method is built upon a theoretical framework that uses as input and output system-averaged quantities, thus being easier to apply in an experimental context where particle resolved information is not available. In this work, we train a deep neural network to predict the self intermediate scattering function of binary mixtures using their static structure factor as input. While its performance is excellent for the temperature range of the training data, the model also retains some transferability in making decent predictions at temperatures lower than the ones it was trained for, or when we use it for similar systems. We also develop an evolutionary strategy that is able to construct a realistic memory function underlying the observed non-Markovian dynamics. This method lets us conclude that the memory function of supercooled liquids can be effectively parameterized as the sum of two stretched exponentials, which physically corresponds to two dominant relaxation modes.

1. Introduction

Understanding the dynamics of supercooled liquids approaching the glass transition represents one of the major challenges in condensed matter science [1–4]. The most striking signature of this phenomenon is the dramatic increase in viscosity or relaxation time upon a relatively mild change in the thermodynamic control parameters. Despite the magnitude of this effect, there are no substantial changes in the microscopic structure of the material, which severely hinders our understanding of the mechanisms underlying the glass transition.

In recent years, machine learning algorithms have been successfully employed to capture subtle changes in the local structure of glassforming materials to create accurate predictors of the dynamics. The first such example is a machine-learned parameter called softness [5–9], which, based on support vector machines and physical intuition, identifies key structural features that strongly correlate with local particle dynamics. Neural networks can also identify local structures [10, 11] and correlate them to local dynamics [12, 13]. Furthermore graph neural networks [14] have shown that the graph structure of each particle's local environment contains significant information to predict its long-time dynamics. It was later demonstrated that more refined observables could be calculated [15] and combined with simpler models to capture the connection between statics and dynamics in glassy systems [16, 17]. Similar results can be achieved even with information theory [18, 19] and dimensionality reduction [19, 20]. Recently, neural networks have also been used to find complex order parameters for glassy dynamics [21]. However all these approaches are based on a

Figure 1. Sketch of our machine learning approach based on averaged static descriptors. The static structure factor is the main input of the model and even though it only shows minute changes with the temperature T, the model was trained to predict the abrupt slowdown from these minimal changes. The DNN explained in section 2.1 predicts $F_s^{\alpha\alpha}(k_p, t)$ from $S^{\alpha\beta}(k)$. Instead with our evolutionary strategy (section 2.2) we are able to construct the memory function $M^{\alpha\beta}(k_p, t)$ that produces the observed dynamics.

particle resolved description of the system, which requires knowledge of the location of every single particle and its precise local environment. Unfortunately these particle resolved quantities are not always easy to measure, and furthermore, single-particle properties do not easily lend themselves to statistical–physical theory development. Hence, a more collective description is often preferred.

Here we propose an alternative approach that is not based on local single-particle features, but instead on system averaged quantities. This approach takes inspiration from collective theories of the glass transition [22–33] that aim to predict the glassiness of the system using collective static observables that do not need to be resolved per particle. The cornerstone of our method consists of rewriting the dynamics of supercooled liquids following the Mori–Zwanzig procedure [22, 34] to obtain a form of a generalized Langevin equation (GLE) called the memory equation [35]. From a mathematical point of view, this equation takes as input the statistically-averaged static structure of the system, mainly through the static structure factor $S(k)$ which is a function of the wave vector k. Using $S(k)$ as the initial boundary condition, the memory equation can then be used to predict the time-dependent dynamics of the system, quantified by the intermediate scattering function $F(k, t)$ at a given time t. The key bottleneck, however, is finding the exact memory function that governs the dynamics of $F(k, t)$; this memory function should account for the dynamical slowdown of supercooled liquids, but its functional form is *a priori* unknown. After decades of intense research, scientists have been able to solve only approximations of this equation, like mode-coupling theory (MCT) [22, 31, 34, 36–43].

In this paper, we use machine-learning to approximate the memory function. In particular we discuss two different approaches to the problem: (i) first we train a deep neural-network (DNN) in order to learn the memory equation using data measured from computer simulations of binary mixtures interacting via Lennard–Jones (LJ) or Weeks–Chandler–Andersen (WCA) potentials. In this approach the DNN plays the role of both the memory equation and the memory function itself. (ii) Then we develop an evolutionary strategy (ES) tailored to identify and construct the memory function that produces the dynamics observed in the simulations.

In figure 1 we sketch the concept of our approach. First we collect extensive simulation data for LJ and WCA binary mixtures. For different values of temperature T and density ρ we measure the static structure factor $S^{\alpha\beta}$ and the self intermediate scattering function $F_s^{\alpha\alpha}$, where α and β are the indexes to represent the species of the mixture. These averaged descriptors are then given to a DNN that we train to predict the intermediate scattering function for a given $S^{\alpha\beta}$, under the assumption introduced in section section 2.1. We show that our DNN achieves excellent performance, thus concluding that the network can learn the memory equation. Next, in order to obtain some physical intuition about the memory function we introduce an ES that, given the intermediate scattering function and the structure of the memory equation, is able to describe the memory function in a parametrized functional form.

Overall, the results of our model are twofold: we introduce an effective approach that is able to rapidly predict the collective dynamics of the system from static measurements, once the model has been trained. Secondly, we propose a representation of the memory function that may be more convenient, and perhaps more realistic, than some state-of-the-art theories. The function parametrized by our machine-learning algorithm can be informative for future efforts aimed at developing a more quantitative theory of the glass transition. In the next sections we will discuss the mechanism of our machine-learning approach, how to generalize it to other systems, and some implications of our findings.

2. Results

2.1. Dynamics of supercooled liquids from neural networks

We train a multi-layer perceptron (MLP) to predict the dynamics of supercooled liquids from static averaged quantities. The MLP is a fully connected DNN with a simple feed-forward architecture. The dynamics is characterized by measuring the intermediate scattering function $F^{\alpha\beta}(k,t)$ from the simulations (details in section 4.1). Rather than a full k-dependent description of the dynamics, we follow the standard procedure [42, 44, 45] for binary mixtures, which consists in the following steps: (i) we focus only on $k = k_{\mathrm{p}} \equiv |\boldsymbol{k}_{\mathrm{peak}}^{AA}|$ which is the location of the main peak of S^{AA} and it is arguably the most descriptive wavenumber for such supercooled mixtures [35], (ii) we consider only the self part of the intermediate scattering function assuming that it also represents the collective dynamics, and (iii) we ignore the mixed term $F^{AB}(k,t)$ since $F_{\mathrm{s}}^{\alpha\beta} = 0$. Hence we end up describing the dynamics using $F_{\mathrm{s}}^{\alpha\alpha}(k_{\mathrm{p}},t)$, where $\alpha = A, B$ represents the two species.

Furthermore, we define $F_{\mathrm{s}}^{\alpha\alpha}(k_{\mathrm{p}},t)$ in the following way:

$$F_{\mathrm{s}}^{\alpha\alpha}(k_{\mathrm{p}},t) = \frac{1}{N}\left\langle \sum_i^N e^{-i\boldsymbol{k}_{\mathrm{p}}\boldsymbol{r}_i^\alpha(0)} e^{i\boldsymbol{k}_{\mathrm{p}}\boldsymbol{r}_i^\alpha(t)} \right\rangle \cdot \underbrace{\frac{1}{N}\left\langle \sum_i^N e^{-i\boldsymbol{k}_{\mathrm{p}}\boldsymbol{r}_i^\alpha(0)} \sum_j^N e^{i\boldsymbol{k}_{\mathrm{p}}\boldsymbol{r}_j^\alpha(0)} \right\rangle}_{\equiv S^{\alpha\alpha}(k_{\mathrm{p}})}, \tag{1}$$

where N is the total number of particles and $\boldsymbol{r}_i^\alpha(t)$ is the position of particle i of species α at time t. Our choice of normalization imposes $F_{\mathrm{s}}^{\alpha\alpha}(k_{\mathrm{p}}, t=0) = S^{\alpha\alpha}(k_{\mathrm{p}})$. In our work, this has two advantages: we can now directly compare our prediction of $F_{\mathrm{s}}^{\alpha\beta}(k_{\mathrm{p}},t)$ with collective theories like MCT under assumptions (i)–(iii), and secondly, our machine learning approach gives more importance to the most glassy (slowest) dynamics, because they are rescaled by a larger factor corresponding to their larger values of $S^{\alpha\beta}(k_{\mathrm{p}})$.

The MLP that we use to predict the dynamics takes as input the static structure factor $S^{\alpha\beta}(k)$ on a uniform grid of $N_k = 100$ wave numbers in the range $0 \leqslant k \leqslant 40\sigma_{AA}$, including the $\alpha \neq \beta$ terms, for a total of 300 values of $S^{\alpha\beta}(k_i)$ (considering the $\alpha\beta = \beta\alpha$ symmetry). In addition the MLP receives the temperature T, the density ρ, a label for the interaction type (i.e. either LJ or WCA), and the logarithm of the time at which it has to predict $F_{\mathrm{s}}^{\alpha\beta}(k_{\mathrm{p}},t)$. With this input–output architecture after training we can easily tune the value of t in the input to reconstruct the full time dependence.

In figure 2 we show that the MLP (detailed in section 4.2) produces good predictions in the full temperature range, from high-temperature liquid to glass on which it was trained. In this figure we also report the results of MCT (dotted lines), which is a theory based on the same static information used as input for the MLP. The results indicate that the MLP significantly outperforms MCT in predicting the realistic dynamics.

To quantify the performance of the MLP we measure its predictivity on data outside of it training set. The model has been trained using 90% of the data available while we use the remaining data to evaluate the R^2 score:

$$R^2 = 1 - \frac{SS_{\mathrm{res}}}{SS_{\mathrm{tot}}}, \tag{2}$$

where SS_{res} is the sum of squared residuals

$$SS_{\mathrm{res}} = \sum_{\mathrm{observation}_i} (\mathrm{prediction}_i - \mathrm{truth}_i)^2 \tag{3}$$

and SS_{tot} is the total sum of squares

$$SS_{\mathrm{tot}} = \sum_{\mathrm{observation}_i} \left(\mathrm{prediction}_i - \overline{\mathrm{prediction}}\right)^2, \tag{4}$$

Figure 2. Predictions of the multi-layer perceptron (MLP) for the self intermediate scattering function of binary mixtures, for the *AA* and *BB* components, normalized as defined in equation (1). We report four state points from the test set (not used for training): the red is a warm liquid, the orange is a supercooled liquid, the yellow is a strongly supercooled liquid and blue is a glass. We compare the MLP (dot-dashed lines) with simulations (solid lines) and mode-coupling theory (dotted line) predictions.

Figure 3. Performance of the multi-layer perceptron (MLP) in predicting the self-intermediate scattering function of binary mixtures. We report the R^2 score as a function of temperature, normalized by T_{mct} which is the temperature at which mode-coupling theory predicts the glass transition. Its value is averaged over states with similar values of T/T_{mct} and the colored region represents the standard deviation of each bin. A value of 1 represents perfect predictions. In the inset we compare the MLP prediction with the target simulation, for a set with $R^2 = 0.76$. We conclude that on average the model predictions are good across the entire temperature range.

where $\bar{..}$ represents the mean. Notice that one observation in equations (3) and (4) corresponds to one time $t = t_i$ of $F_s(k_p, t)$ for any given T and ρ, so the sum runs over all the temperatures T, densities ρ and times t_i. Moreover, the train/test split is performed separating 10% of the {LJ/WCA; T; ρ} states rather than 10% of all the data points, in order to avoid training the MLP using data strongly correlated with the test set.

A value of $R^2 = 1$ corresponds to a perfect fit, so R^2 is an effective measure to evaluate the quality of the model. In figure 3 we report the R^2 score as a function of T/T_{mct}, where T_{mct} is the temperature at which fit-parameter-free MCT predicts the glass transition, reported in table 1. With this normalization we can average simulations at different densities and different pairwise interactions to produce a single curve. In this work we are not interested in the precise estimation of the critical point of MCT for our mixtures, but a more detailed discussion is available in [3, 44, 46]. In the supplementary information we show the root mean squared error (RMSE) for the same data. The trend of the curve shows that the performance of the MLP starts to drop at $T < T_{mct}$, when the two-step relaxation becomes more prominent and thus the intermediate scattering function is a more 'complex' function with more features to predict. At $T < 0.6T_{mct}$ the MLP

Table 1. Temperatures that we use to normalize the data. The values of $k_B T_{mct}/\epsilon_{AA}$ correspond to the critical temperature of mode-coupling theory for the WCA and LJ mixtures [3, 44, 46], below which fit-parameter-free MCT predicts that the system is a glass.

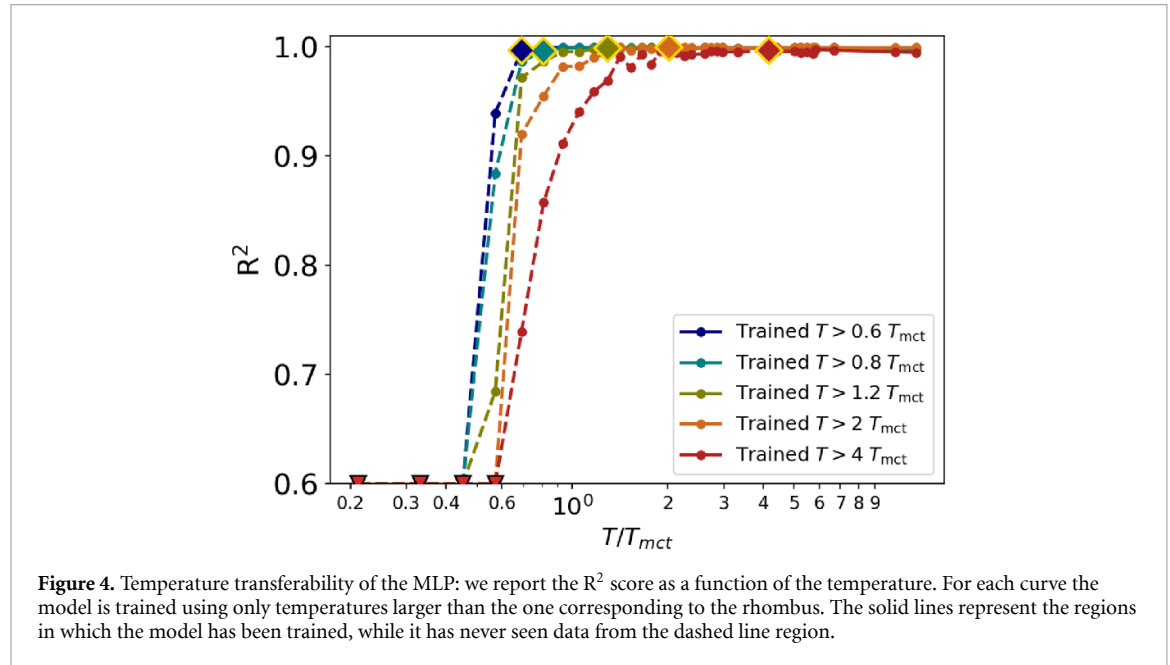| $\frac{k_B T_{mct}}{\epsilon_{AA}}$	$\rho\sigma_{AA}^3 = 1.2$	$\rho\sigma_{AA}^3 = 1.4$	$\rho\sigma_{AA}^3 = 1.6$	$\rho\sigma_{AA}^3 = 1.8$
WCA	0.74	1.77	3.49	5.10
LJ	0.90	1.87	3.53	5.10

Figure 4. Temperature transferability of the MLP: we report the R^2 score as a function of the temperature. For each curve the model is trained using only temperatures larger than the one corresponding to the rhombus. The solid lines represent the regions in which the model has been trained, while it has never seen data from the dashed line region.

predictions exhibit pronounced fluctuations, but the predictions are still good on average, as we highlight in the inset of figure 3. Surprisingly at even lower temperatures, when the dynamics is even slower, the performances increase, approaching again the perfect $R^2 = 1$ score. This is a consequence of the fact that in the glass the second relaxation happens at $t > t_{max}$, which is outside the time window that the model observes, so F is characterized by a single relaxation, thus being easier to learn for the model.

2.1.1. Transferability

2.1.1.1. Temperature transferability

One of the main strengths of machine learning is the possibility to train a model in a more favorable situation (e.g. when there is more data available) and deploy it in a less favorable one. For computer simulations the hardest region to sample is the low temperature regime, because the dynamics is much slower and more time is required to sample all relevant structural rearrangements. This means that it is much easier to collect data at high T than at low T.

Here we show the temperature transferability of our model, i.e. its performance at temperatures different from the training. We report in figure 4 the R^2 score when the MLP is trained only using data for $T > T_0$. The RMSE is reported in the supplementary information. As expected the model performs excellently in its training region (solid lines), but outside (dashed lines) its score starts dropping. The results of figure 4 suggest the existence of two regimes: (i) for $T > 0.8 T_{mct}$ we have full transferability. It is in fact possible to train the model at high temperature and retain good predictions, as evidenced by the red curve in figure 4. (ii) When $T < 0.8 T_{mct}$ the transferability is restricted. In figure 4 we show that the best we can do is to transfer the training at $T > 1.2 T_{mct}$ down to $T > 0.5 T_{mct}$.

Overall we can conclude that our approach is transferable in the (i) MCT regime corresponding to $T > 0.8 T_{mct}$, while the transferability is very limited in the (ii) $T < 0.8 T_{mct}$ activated regime. This low temperature region is in fact characterized by the appearance of heterogeneous activated dynamics and facilitation, that become dominant at low temperature [47, 48]. We hypothesize that this different relaxation mechanism is the reason behind the lack of transferability. In summary, except for the very low temperature region, our approach presents some degree of temperature transferability that we can use to reduce the need for data that are harder to collect.
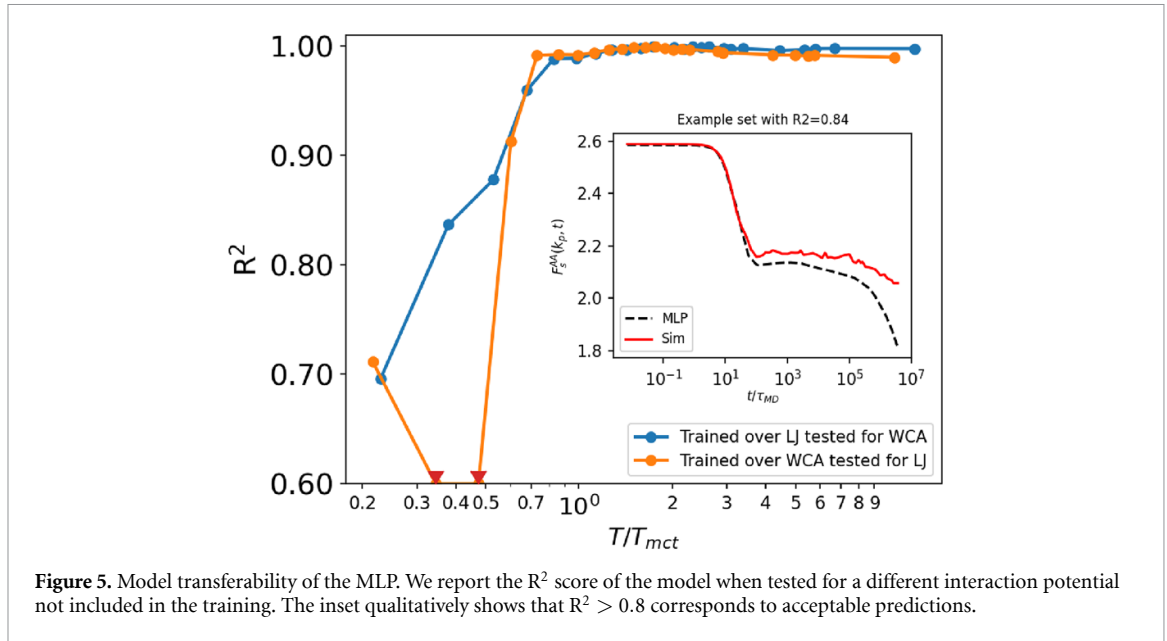
Figure 5. Model transferability of the MLP. We report the R^2 score of the model when tested for a different interaction potential not included in the training. The inset qualitatively shows that R$^2 > 0.8$ corresponds to acceptable predictions.

2.1.1.2. Model transferability

We show in this section the performance of the MLP in predicting the dynamics of computer simulations produced with a different interaction than the one the MLP has been trained on. Our library consists of simulation data for binary LJ and binary WCA, so their dynamics is similar at high temperature and/or high density, but it becomes significantly different approaching T_{mct}. In figure 5 we show the prediction of the MLP when it is trained over LJ and tested over WCA and vice versa, reporting the R^2 score as a function of normalized temperature. We see that at $T > 0.6T_{\text{mct}}$ the MLP produces excellent predictions even when it is trained for a different interaction potential. However the quality drops for very low temperature, when minute differences between the LJ and WCA structures are amplified to enormous differences in dynamics [31, 43, 44, 49–51] and configurational entropy [52–54]. In particular the model performs poorly when it is trained using the WCA potential and tested over the LJ data. This is a consequence of the fact that the LJ model becomes a glass at a higher temperature compared to the WCA model [31, 44, 49], so there is a region where the LJ model is infinitely slower than the WCA.

Overall we see that it is possible to use data measured from another system to obtain reliable predictions for a different (but relatively similar) system. However those predictions become unreliable at very low T, when approaching the glass transition.
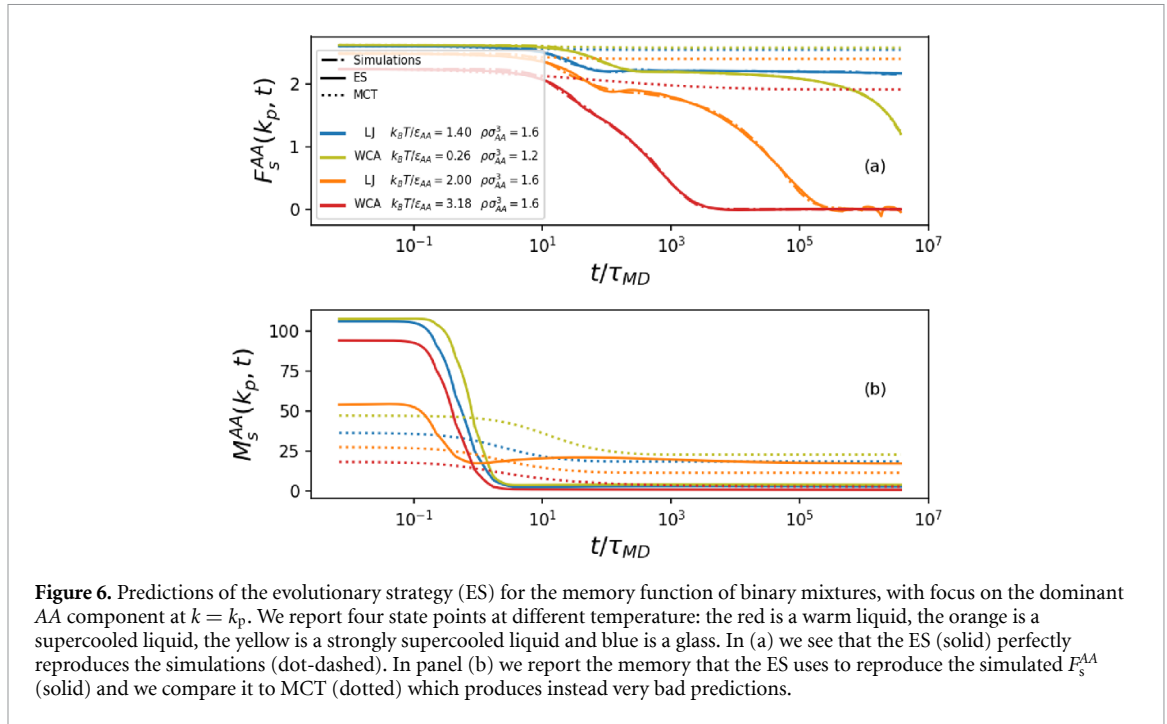
2.2. ES for the memory function

In the previous section we have seen how to numerically correlate static structural information with the dynamics of the system. Such a deep learning approach however does not tell us much about the physics of the system. Here we employ a physics-informed strategy that is created in order to avoid the application of a *black box* [55, 56], but instead we bound the machine learning model to play the role of a single physical unknown function: the memory function.

In order to develop this physics-informed strategy we start from the exact memory equation that describes the overdamped dynamics of liquids [35]:

$$\frac{\mathrm{d}F_s^{\alpha\beta}(k_\mathrm{p}, t)}{\mathrm{d}t} + \Omega^2_{\alpha\beta}(k_\mathrm{p})F_s^{\alpha\beta}(k_\mathrm{p}, t) + \int_0^t \mathrm{d}\tau M^{\alpha\gamma}(k_\mathrm{p}, t-\tau)\frac{\partial F_s^{\gamma\beta}(k_\mathrm{p}, \tau)}{\partial\tau} = 0, \tag{5}$$

where $\Omega^2_{\alpha\beta}(k_\mathrm{p})$ is a constant representing the vibrational term [22, 34]. While the equation above is formally exact, unfortunately the memory function $M^{\alpha\beta}(k_\mathrm{p}, t-\tau)$ that appears in the integral is unknown, and thus the equation cannot be solved. MCT is based on an uncontrolled approximation of this memory function that lead to excellent semi-quantitative results [34, 41, 57], but its predictions are not able to exactly capture the full phenomenology of the glass transition [44]. While there are some ways to invert equation (5) [51, 58], and various approaches to numerically estimate the memory function [59–62], there is no consensus on the general form of M. More importantly, most of these procedures to calculate M are very sensitive to noise, thus requiring refined data, and overall they are often computationally expensive.

Figure 6. Predictions of the evolutionary strategy (ES) for the memory function of binary mixtures, with focus on the dominant *AA* component at $k = k_\mathrm{p}$. We report four state points at different temperature: the red is a warm liquid, the orange is a supercooled liquid, the yellow is a strongly supercooled liquid and blue is a glass. In (a) we see that the ES (solid) perfectly reproduces the simulations (dot-dashed). In panel (b) we report the memory that the ES uses to reproduce the simulated F_s^{AA} (solid) and we compare it to MCT (dotted) which produces instead very bad predictions.

Here we show that it is possible to effectively parameterize M as a sum of stretched exponentials, whose coefficients can be determined by an evolutionary algorithm. In choosing this exponential representation, we have drawn inspiration from theories such as MCT [22, 63] for which the known (albeit approximate) memory function typically has a similar structure as the intermediate scattering function, combined with the fact that $F_\mathrm{s}(k,t)$ is known experimentally to behave as a stretched exponential for long times [64]. Thus, we represent the memory function in the following way:

$$M^{\alpha\beta}(k,t) = c^{\alpha\beta}(k) + \sum_{i=0}^{N_{\exp}} w_i^{\alpha\beta}(k) \times \left\{ \left[1 - a_i^{\alpha\beta}(k) \right] e^{ -\left(\frac{t}{\tau_i^{\alpha\beta}(k)} \right)^{b_i^{\alpha\beta}(k)} } + a_i^{\alpha\beta}(k) \right\} \tag{6}$$
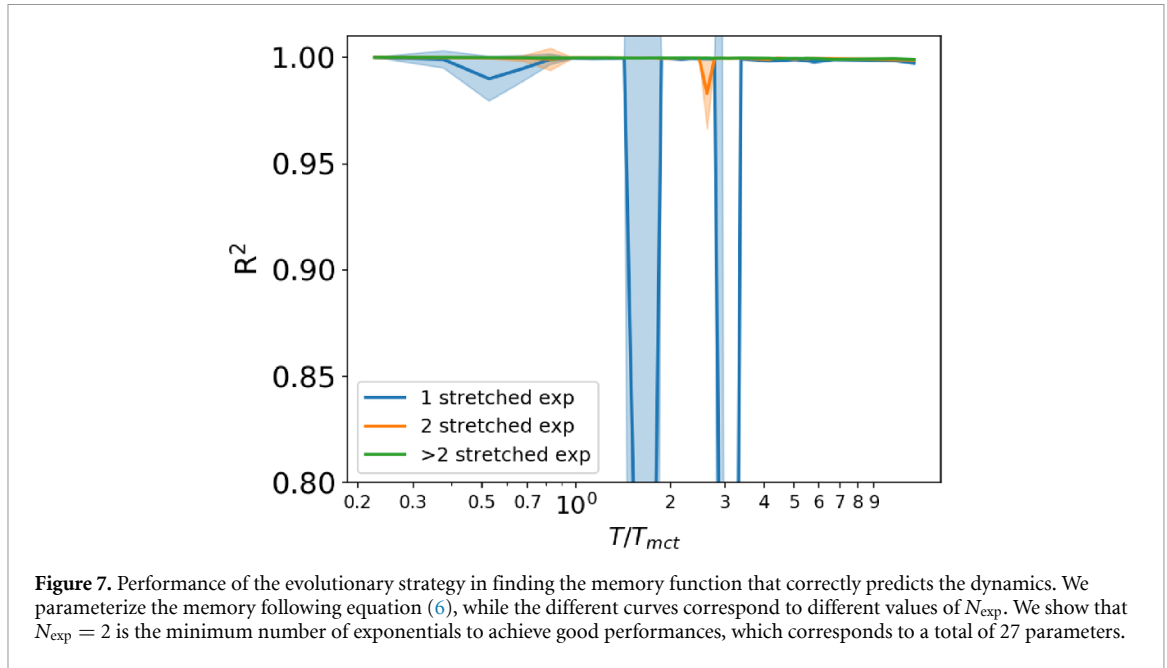
which is a sum of $N_{\exp}$ stretched exponentials, for a total of $3(1 + 4N_{\exp})$ parameters for every state. Notice that our parametrization also allows for standard or compressed exponentials, according to the value of $b_i^{\alpha\beta}(k)$. As previously discussed (section 2.1), we focus only on $k = k_{\mathrm{peak}}^{AA}$, which is usually the most important wavevector in simple glassformers such as the systems studied here.

The details of our ES are discussed in section 4.4. Briefly, the ES is an optimization technique inspired by evolution and natural selection. It works by creating new generations of memory functions (parameterized as equation (6)) by cross-breeding the previous generation and adding random mutations. For each of them we then solve the memory equation (equation (5)) obtaining $\{F_{\mathrm{ES}}^{\alpha\alpha}(t_i)\}$ and then we do one step of evolution favoring the individuals that minimize the following loss function:

$$\mathcal{L}_{\mathrm{ES}} = \sum_{t_i, \alpha} [F_\mathrm{s}^{\alpha\alpha}(k_\mathrm{P}, t_i) - F_{\mathrm{ES}}^{\alpha\alpha}(t_i)]^2 + \text{regularization} \tag{7}$$

which is the squared difference between predicted and observed dynamics. In the supplementary information we verify that the ES solutions are physically reasonable by replacing the intermediate scattering function with the MCT approximation $F_\mathrm{s}^{\alpha\alpha}(k_\mathrm{P}, t_i) \rightarrow F_{\mathrm{MCT}}^{\alpha\alpha}(k_\mathrm{P}, t_i)$ and confirming that the ES converges towards $M_{\mathrm{MCT}}^{\alpha\beta}(k_p, t)$. Notice however that equation (6) allows unrealistic short time non-monotonic behavior, that we observe sometimes. Overall, we believe that our ES solutions still retain a fair degree of realism.

The results of the ES are reported in figure 6. In panel (a) we show that the ES is able to recover the simulated self intermediate scattering function $F_\mathrm{s}^{\alpha\alpha}(k_\mathrm{P}, t_i)$ that we are targeting. The figure presents four curves that range from the liquid to the glass phase. We have also verified that the ES solution converges to the simulated $F_\mathrm{s}^{\alpha\alpha}(k_\mathrm{p}, t)$ for any temperature and density that we have available. Even though we have only focused on $k = k_\mathrm{p}$, we speculate that our approach should work for any value of k since the structure of equations (5) and (6) does not depend on k. Note however that, unlike MCT, this parametrized memory

Figure 7. Performance of the evolutionary strategy in finding the memory function that correctly predicts the dynamics. We parameterize the memory following equation (6), while the different curves correspond to different values of N_{exp}. We show that $N_{\text{exp}} = 2$ is the minimum number of exponentials to achieve good performances, which corresponds to a total of 27 parameters.

function does not contain explicit couplings among different wave numbers, nor any self-consistent feedback mechanism to drive dynamical slowdown.

In figure 6(b) we report the results for the memory function. Following the parametrization introduced in equation (6) the ES is able to converge to memory functions that produce the realistic $F_S^{\alpha\alpha}(k_{\text{p}}, t)$ of figure 6(a). In particular the memory functions presented in figure 6(b) are obtained setting $N_{\text{exp}} = 2$. We highlight the differences between the M at which the ES converged, with the MCT approximation (dashed lines), that unsurprisingly overestimates the glassiness of these systems [35, 42, 44]. We conclude that the ES method we introduced is an effective way to obtain a realistic memory function.

Lastly we discuss the physical insights that we can harness from our evolutionary approach. We have seen that the ES converges to the simulated intermediate scattering function if we model the memory function following equation (6). Since equation (5) is exact, we know that the real memory function has to produce the same output as our ES when propagated through equation (5). This means that we can look at the structure of our ES solutions to have some intuition about the real memory. In particular, in figure 7 we report the R^2 for different parametrizations containing a different number of stretched exponentials and we verify that a good solution needs at least $N_{\text{exp}} = 2$. In the supplementary information we also show that the second stretched exponential becomes relevant only approaching the glass transition, in the same range where localized unstable modes become relevant [65], and when energy-driven and entropy-driven activation start to compete [66]. We also know that activation and facilitation become relevant only at very low temperature [47, 48]. This suggests that there are at least two separate relaxation channels that real materials follow while relaxing, with the second channel becoming dominant below T_{mct}, i.e. the temperature that is often identified with a crossover [67]. It is important to recall however that the T_{mct} used in our work is obtained from fit-parameter-free MCT, while other works have considered different definitions of T_{mct}, and hence these comparisons should be treated with caution.

Moreover, the results derived from the ES may also provide cues for analytical modeling. Models like MCT invoke uncontrolled approximations to solve equation (5). One possibility is the exponential closure that assumes that $M^{\alpha\beta}(k, t) \sim \exp\left[-\Omega_{\alpha\beta}^2(k)t\right]$. This schematic model is simple enough to be solved analytically, but its results are not satisfying [30, 63, 68–70]. Inspired by our ES we may propose a double stretched-exponential memory defined as

$$M^{\alpha\beta}(k, t) = \exp\left[-\Omega_{\alpha\beta}^2(k)t^{\nu}\right] + K^{\alpha\beta}(k)\exp\left[-\tilde{\Omega}_{\alpha\beta}^2(k)t^{\mu}\right]. \tag{8}$$

This schematic model retains a structure of equation (5) similar to the exponential closure, but according to our ES results it should be more appropriate to describe the glassy dynamics, provided that it is properly parametrized. Note however that the correct parametrization would still require numerical fitting (e.g. via an ES). Nonetheless, we argue that our ES strategy can be used to explore different functional forms that may improve upon the conventional MCT closure approximation.

In summary we have defined here a simple approach to determine the memory function M given the self-intermediate scattering function F_s. The ES is fast and reliable across all the temperatures and densities studied. The results suggest that at least two relaxation channels need to be considered, giving rise to the machine-learning-inspired schematic model of equation (8).

3. Discussion

The goal of this study was to show how a neural network can understand and predict the dynamics of supercooled liquids from static information. Our approach is based on averaged quantities such as static structure factors that are easier to measure experimentally, compared to particle resolved ones. We are also able to interpret the mechanism of the machine learning model to gain some physical intuition about the glass transition.

In figures 2 and 3 we show that we can train a MLP to efficiently predict the dynamics (represented by the self intermediate scattering function $F_s^{\alpha\beta}$), using only statistically-averaged static information. This implies that dynamical information is encoded in the static structure of the system, similarly to how higher-order correlations are partially encoded into two body structure [50]. Furthermore since the main input of our neural network is the static structure factor $S^{\alpha\beta}$, our results corroborate the idea that two body static correlations when elaborated using an expressive approach like our MLP, are enough to describe the dynamics.

One of the main problems in studying systems close to the glass transition is data collection, because experiments and simulations at deeply supercooled temperatures are slow. We show that our MLP performs relatively well when only high temperature data are provided for training (figure 4) or when the MLP is used to make predictions for different variations of the model that is used for training (figure 5). Unfortunately this transferability drops when the system is approaching the glass transition, because we know that minute changes in the structure correspond to enormous changes in the dynamics. Overall this means that, if the interest is in the liquid regime it is possible to fully exploit the MLP transferability by training the model where data is easily available, but to accurately describe the glassy regime, glassy data is actually required.

We then develop a physics-inspired method to obtain an expression for the memory function that realistically describe our data. Instead of using deep learning as a black box to connect statics to dynamics, we rewrite the dynamics as a memory function and we replace this memory with our machine learning model. Our approach circumvents inverse Laplace transforms, which can be computationally expensive, by using an ES that parametrizes a pre-defined functional form for the memory function. We show in figure 6 that the ES easily converges to memory functions that reproduce the real dynamics observed in simulations.

Our physics-inspired evolutionary approach is also able to give some intuition about the physics. The results in figure 7 let us conclude that the memory function can be effectively parameterized as a sum of two stretched exponentials. We can interpret those two stretched exponentials as two relaxation processes that describe the complex multiscale relaxation of the glassy liquid, and we can also see that only one is needed to describe the liquid phase (SI figure S7). One possible interpretation for the existence of two dominant relaxation channels might be dynamic heterogeneity, i.e. the coexistence of transiently fast and slow groups of particles, which is known to emerge at temperatures below T_{mct} [6, 12, 47]. However, more work is needed to identify the microscopic physical origins of the two stretched exponentials and to pinpoint the temperature regime where this effect is relevant. This intuition motivates us to explore a schematic model where the memory function is exactly represented as two stretched exponentials, but we leave that for future work.

In conclusion, we have introduced two data-driven tools to evaluate and describe the dynamics of supercooled liquids. The neural network that we propose can be efficiently trained and deployed to predict the self-intermediate scattering function from averaged quantities which are simple to measure experimentally. Lastly we have discussed a way to obtain an effective memory function using an ES, concluding that the memory can be reasonably represented as a sum of two stretched exponentials. We believe that our machine-learning method, once trained, can be efficiently applied to predict the dynamics of many other glass forming mixtures and that data-driven approaches to find suitable functional forms of the memory function may help guide the development of more effective theories to describe the glass transition.

4. Methods

4.1. Computer simulations
The models reported in this paper have been trained and tested using simulation data of two binary mixtures: the Kob–Andersen binary LJ mixture [71] and its WCA truncation [72]. They are three-dimensional 80 : 20 mixtures of particles $A : B$ interacting with each other via

$$V_{\alpha\beta}(r) = \begin{cases} 4\epsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r}\right)^6 + C_{\alpha\beta}\right], & r \leqslant r_{\alpha\beta}^c \\ \\ \hspace{4cm} , & r > r_{\alpha\beta}^c \end{cases} \tag{9}$$

where the cutoff radius $r_{\alpha\beta}^c$ is $2.5\sigma_{\alpha\beta}$ for LJ, and $r_{\alpha\beta}^c$ is $2^{1/6}\sigma_{\alpha\beta}$ for WCA [72]. The choice of $C_{\alpha\beta}$ secures that $V_{\alpha\beta}(r_{\alpha\beta}^c) = 0$. As usual [71], we set $\epsilon_{AB}/\epsilon_{AA} = 1.5, \epsilon_{BB}/\epsilon_{AA} = 0.5, \sigma_{AB}/\sigma_{AA} = 0.8, \sigma_{BB}/\sigma_{AA} = 0.88$.

For these mixtures we perform molecular dynamics simulations in the *NVE* ensemble using HOOMD-blue [73]. First we equilibrate the system at different densities $\rho\sigma_{AA}^3 \in [1.2, 1.8]$ and temperatures $k_B T/\epsilon_{AA} \in [0.2, 15]$. We impose periodic boundary conditions and set the box at $L = 10\sigma_{AA}$ while tuning the density by changing the number of particles $N \in [1200, 2000]$. All the particles have the same mass m. We run the simulations for 10^8 timesteps of size $dt = 10^{-3}\tau_{MD}$ with $\tau_{MD} = \left(m\sigma_{AA}^2/48\epsilon_{AA}\right)^{1/2}$, which allows us to sample up to $t_{max} = 3 \times 10^6 \tau_{MD}$. Notice that the trajectories that are deeply in the glass regime (i.e. where $F_s^{\alpha\alpha}(k, t \to \infty) > 0$) cannot be fully equilibrated due to the ergodicity breaking that defines the glass transition. Additionally, even though the mixture is designed to avoid crystallization, it is still possible for some specific trajectories to crystallize; in that case we remove such occurrences from our data. Finally, we use the simulated trajectories to calculate the partial static structure factors $S^{\alpha\beta}(k)$ and the self intermediate scattering functions $F_s^{\alpha\beta}(k, t)$.

4.2. MLP

We train a MLP to predict the dynamics of supercooled mixtures from static information. The results reported in section 2.1 are produced by a MLP. The network consists of five hidden layers of size $\{500, 400, 400, 300, 200\}$, interposed by ReLU activation functions, that transform the 304 input features $\left\{S^{AA}(k_1), \ldots, S^{BB}(k_{100}), T, \rho, \text{interaction}, t\right\}$ into the output $\left\{F_s^{AA}(k_p, t), F_s^{BB}(k_p, t)\right\}$. The MLP is trained for 1000 iterations, taking approximately one hour on a standard Intel i7 CPU. We also use $L1$ loss and Lasso regularization [74] with the adam optimization algorithm [75].

4.3. Memory equation and MCT

In this paper we numerically solve equation (5), which is called the memory equation [35]. This is an equation of the same class as the GLE and we believe that out method can be tailored to solve a wide category of GLE equations with a structure similar to equation (5), using physical intuition.

In general, it is possible to exactly describe the dynamics of liquids by deriving an expression for

$$F^{\alpha\beta}(k, t) = \frac{1}{N}\left\langle \sum_i^N e^{-ik r_i^\alpha(0)} \sum_j^N e^{ik r_j^\beta(t)} \right\rangle, \tag{10}$$

from the solution of the overdamped equation

$$\frac{dF^{\alpha\beta}(k, t)}{dt} + \Omega_{\alpha\beta}^2(k)F^{\alpha\beta}(k, t) + \int_0^t d\tau M^{\alpha\beta}(k, t - \tau)\frac{\partial F_s^{\alpha\beta}(k, \tau)}{\partial \tau} = 0, \tag{11}$$

but unfortunately the memory function $M^{\alpha\beta}(k, t)$ is unknown. Notice that in this paper we are mainly interested in solving equation (5) which corresponds to equation (11) if $F^{\alpha\beta}(k, t)$ is replaced with $F_s^{\alpha\alpha}(k_p, t)$. We will also assume that the memory function is the same in equations (5) and (11).

The memory equation can only be solved using some approximations like MCT [22, 31, 34, 37, 38, 43], for this reason we also refer to the MCT equation. The MCT approximation applied to equation (5) consists of the following definition of the memory function:

$$M_{mct}^{\alpha\beta}(k, t) = \frac{1}{2k^2}\frac{\rho}{x_\alpha x_\beta}\sum_{\substack{\alpha'\beta'\\\alpha''\beta''}}\int\frac{d^3\boldsymbol{q}}{(2\pi)^2}\cdot$$
$$\cdot V_{\alpha\alpha'\alpha''}(\boldsymbol{q}, \boldsymbol{k}, \boldsymbol{k} - \boldsymbol{q})F_s^{\alpha'\beta'}(k, t)\cdot$$
$$\cdot F_s^{\alpha''\beta''}(k, t)V_{\beta\beta'\beta''}(\boldsymbol{q}, \boldsymbol{k}, \boldsymbol{k} - \boldsymbol{q}), \tag{12}$$

where $x_\alpha = N_\alpha/N$ is the density of species α and the vertex function corresponds to

$$V_{\alpha\beta\gamma}(\boldsymbol{k}, \boldsymbol{q}, \boldsymbol{p}) = \left(\hat{k}\cdot\boldsymbol{q}\right)c^{\alpha\beta}(q)\delta_{\alpha\beta} + \left(\hat{k}\cdot\boldsymbol{p}\right)c^{\alpha\gamma}(p)\delta_{\alpha\gamma}, \tag{13}$$

with $S_{\alpha\beta}^{-1}(k) = \delta_{\alpha\beta}/x_\alpha - \rho c_{\alpha\beta}(k)$.

So overall, the inputs for the MCT equation are the bulk density ρ, the temperature T, and the structure factor $S^{\alpha\beta}(k)$. In our numerical solution of the MCT equation we follow all the steps discussed in [31, 41, 43, 45] over a grid of $N_k = 100$ points. As an alternative to the MCT approximation, we also solve equation (5) using equation (6) instead of $M_{mct}^{\alpha\beta}$, from the same inputs. In summary, at any given (T, ρ) we only require $S^{\alpha\beta}(k)$ as input to predict the microscopic relaxation dynamics of the system, either with MCT or with our parametrization of the memory identified by the ES.

4.4. ES

ES is a class of machine learning optimization algorithms that are inspired by natural evolution in the following way: at every iteration (or generation), a population of parameters (the genotypes) are perturbed by cross-breeding and mutations and their objective function (fitness) is evaluated. Then the highest scoring parameters are recombined to populate the next generation, iteratively until the objective function is optimized. The huge advantage of this class of algorithms is that they do not require back-propagation, which is particularly useful when the objective function is a complex integro-differential equation like our GLE (equation (5)).

We use covariance matrix adaptation evolution strategy [76] (CMA-ES), a widely known method of the ES class which describes the population by a multivariate Gaussian. The algorithm is available as a python package [77]. Another advantage is that it does not require any hyper-parameter except for the initial condition and the population size represented by the initial variance of the gaussian σ. We then use CMA-ES to find the best function $M_{ES}^{\alpha\beta}$ parametrized as equation (6) that reproduces the real dynamics. In practice at each step we propagate $M_{ES}^{\alpha\beta}$ through equation (5) and we compare its output $F_{ES}^{\alpha\alpha}(t)$ to the real dynamics $F_S^{\alpha\alpha}(k_p, t)$, evaluating the fitness function $\mathcal{F} = -\mathcal{L}$ (equation (7)) for the evolution. We use as initial condition $M = M_{mct}$ and for the initial population we set $\sigma = 5 \times 10^{-4}$. We use Lasso regularization to evaluate equation (7). Usually the procedure converges below an error $\epsilon < 10^{-6}$ within a couple of hours of evolution. Results in figure 7 show that the F predicted by ES is very close to the real dynamics, so we conclude that M is well represented by two stretched exponentials.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

Author contributions

S C performed the simulations. S C developed the MLP model. S C, M C, E B developed the ES. S C wrote the paper with the help of M C, E B, M D and L M C J. M D, L M C J provided resources and supervision.

ORCID iDs

Simone Ciarella ● https://orcid.org/0000-0002-9247-139X
Marjolein Dijkstra ● https://orcid.org/0000-0002-9166-6478
Liesbeth M C Janssen ● https://orcid.org/0000-0001-5283-1330

References

[1] Anderson P W 1995 Through the glass lightly *Science* **267** 1615–6
[2] Lubchenko V and Wolynes P G 2007 Theory of structural glasses and supercooled liquids *Annu. Rev. Phys. Chem.* **58** 235–66
[3] Berthier L and Biroli G 2011 Theoretical perspective on the glass transition and amorphous materials *Rev. Mod. Phys.* **83** 587–645
[4] Langer J S 2014 Theories of glass formation and the glass transition *Rep. Prog. Phys.* **77** 042501
[5] Cubuk E D, Schoenholz S S, Rieser J M, Malone B D, Rottler J, Durian D J, Kaxiras E and Liu A J 2015 Identifying structural flow defects in disordered solids using machine-learning methods *Phys. Rev. Lett.* **114** 108001
[6] Schoenholz S S, Cubuk E D, Sussman D M, Kaxiras E and Liu A J 2016 A structural approach to relaxation in glassy liquids *Nat. Phys.* **12** 469–71
[7] Ma X, Davidson Z S, Still T, Ivancic R J S, Schoenholz S, Liu A and Yodh A 2019 Heterogeneous activation, local structure and softness in supercooled colloidal liquids *Phys. Rev. Lett.* **122** 028001

[8] Cubuk E D, Liu A J, Kaxiras E and Schoenholz S S 2020 Unifying framework for strong and fragile liquids via machine learning: a study of liquid silica (arXiv:2008.09681)

[9] Tah I, Ridout S A and Liu A J 2022 Fragility in glassy liquids: a structural approach based on machine learning *J. Chem. Phys.* **157** 124501

[10] Boattini E, Ram M, Smallenburg F and Filion L 2018 Neural-network-based order parameters for classification of binary hard-sphere crystal structures *Mol. Phys.* **116** 3066–75

[11] Boattini E, Dijkstra M and Filion L 2019 Unsupervised learning for local structure detection in colloidal systems *J. Chem. Phys.* **151** 154901

[12] Boattini E, Marín-Aguilar S, Mitra S, Foffi G, Smallenburg F and Filion L 2020 Autonomously revealing hidden local structures in supercooled liquids *Nat. Commun.* **11** 5479

[13] Oyama N, Koyama S and Kawasaki T 2022 What do deep neural networks find in disordered structures of glasses? (arXiv:2208.00349)

[14] Bapst V *et al* 2020 Unveiling the predictive power of static structure in glassy systems *Nat. Phys.* **16** 448–54

[15] Boattini E, Smallenburg F and Filion L 2021 Averaging local structure to predict the dynamic propensity in supercooled liquids *Phys. Rev. Lett.* **127** 088007

[16] Alkemade R M, Boattini E, Filion L and Smallenburg F 2022 Comparing machine learning techniques for predicting glassy dynamics *J. Chem. Phys.* **156** 204503

[17] Jung G, Biroli G and Berthier L 2022 Predicting dynamic heterogeneity in glass-forming liquids by physics-informed machine learning (arXiv:2210.16623)

[18] Jack R L, Dunleavy A J and Royall C P 2014 Information-theoretic measurements of coupling between structure and dynamics in glass formers *Phys. Rev. Lett.* **113** 095703

[19] Paret J, Jack R L and Coslovich D 2020 Assessing the structural heterogeneity of supercooled liquids through community inference *J. Chem. Phys.* **152** 144502

[20] Coslovich D, Jack R L and Paret J 2022 Dimensionality reduction of local structure in glassy binary mixtures *J. Chem. Phys.* **157** 204503

[21] Zhang K, Li X, Jin Y and Jiang Y 2022 Machine learning glass caging order parameters with an artificial nested neural network *Soft Matter* **18** 6270–7

[22] Götze W and Sjogren L 1992 Relaxation processes in supercooled liquids *Rep. Prog. Phys.* **55** 241–376

[23] Xia X and Wolynes P G 2000 Fragilities of liquids predicted from the random first order transition theory of glasses *Proc. Natl Acad. Sci.* **97** 2990–4

[24] Tarjus G, Kivelson S A, Nussinov Z and Viot P 2005 The frustration-based approach of supercooled liquids and the glass transition: a review and critical assessment *J. Phys.: Condens. Matter* **17** R1143–82

[25] Sausset F, Tarjus G and Viot P 2008 Tuning the fragility of a glass-forming liquid by curving space *Phys. Rev. Lett.* **101** 155701

[26] Biroli G and Bouchaud J-P 2012 The random first-order transition theory of glasses: a critical assessment *Structural Glasses and Supercooled Liquids: Theory, Experiment, and Applications* ed P G Wolynes and V Lubchenko ch 2 31–113

[27] Dell Z E and Schweizer K S 2015 Microscopic theory for the role of attractive forces in the dynamics of supercooled liquids *Phys. Rev. Lett.* **115** 205702

[28] Rizzo T and Voigtmann T 2015 Qualitative features at the glass crossover *Europhys. Lett.* **111** 56008

[29] Rizzo T and Voigtmann T 2020 Solvable models of supercooled liquids in three dimensions *Phys. Rev. Lett.* **124** 195501

[30] Janssen L M C and Reichman D R 2015 Microscopic dynamics of supercooled liquids from first principles *Phys. Rev. Lett.* **115** 205701

[31] Ciarella S, Luo C, Debets V E and Janssen L M C 2021 Multi-component generalized mode-coupling theory: predicting dynamics from structure in glassy mixtures *Eur. Phys. J.* E **44** 91

[32] Liu C, Biroli G, Reichman D R and Szamel G 2021 Dynamics of liquids in the large-dimensional limit *Phys. Rev.* E **104** 054606

[33] Szamel G 2022 An alternative, dynamic density functional-like theory for time-dependent density fluctuations in glass-forming fluids *J. Chem. Phys.* **156** 191102

[34] Reichman D R and Charbonneau P 2005 Mode-coupling theory *J. Stat. Mech.* P05013

[35] Kob W 2003 Course 5: supercooled liquids, the glass transition, and computer simulations *Slow Relaxations and Nonequilibrium Dynamics in Condensed Matter* (*Les Houches Session LXXVII, 1–26 July 2002*) (Berlin: Springer) pp 199–269

[36] Franosch T, Fuchs M, Götze W, Mayr M R and Singh A P 1997 Asymptotic laws and preasymptotic correction formulas for the relaxation near glass-transition singularities *Phys. Rev.* E **55** 7153–76

[37] Voigtmann T 2003 Mode coupling theory of the glass transition in binary mixtures (available at: https://mediatum.ub.tum.de/603008)

[38] Götze W 2008 *Complex Dynamics of Glass-Forming Liquids: A Mode-Coupling Theory* (Oxford: Oxford University Press)

[39] Szamel G 2003 Colloidal glass transition: beyond mode-coupling theory *Phys. Rev. Lett.* **90** 228301

[40] Brader J M, Voigtmann T, Fuchs M, Larson R G and Cates M E 2009 Glass rheology: from mode-coupling theory to a dynamical yield criterion *Proc. Natl Acad. Sci.* **106** 15186–91

[41] Weysser F, Puertas A M, Fuchs M and Voigtmann T 2010 Structural relaxation of polydisperse hard spheres: comparison of the mode-coupling theory to a Langevin dynamics simulation *Phys. Rev.* E **82** 011504

[42] Janssen L M C 2018 Mode-coupling theory of the glass transition: a primer *Front. Phys.* **6** 97

[43] Debets V E, Luo C, Ciarella S and Janssen L M C 2021 Generalized mode-coupling theory for mixtures of Brownian particles *Phys. Rev.* E **104** 065302

[44] Berthier L and Tarjus G 2010 Critical test of the mode-coupling theory of the glass transition *Phys. Rev.* E **82** 031502

[45] Flenner E and Szamel G 2005 Relaxation in a glassy binary mixture: comparison of the mode-coupling theory to a Brownian dynamics simulation *Phys. Rev.* E **72** 031508

[46] Voigtmann T 2008 Idealized glass transitions under pressure: dynamics versus thermodynamics *Phys. Rev. Lett.* **101** 1–4

[47] Guiselin B, Scalliet C and Berthier L 2022 Microscopic origin of excess wings in relaxation spectra of supercooled liquids *Nat. Phys.* **18** 468–72

[48] Scalliet C, Guiselin B and Berthier L 2022 Thirty milliseconds in the life of a supercooled liquid (arXiv:2207.00491)

[49] Berthier L and Tarjus G 2011 The role of attractive forces in viscous liquids *J. Chem. Phys.* **134** 214503

[50] Coslovich D 2013 Static triplet correlations in glass-forming liquids: a molecular dynamics study *J. Chem. Phys.* **138** 12A539

[51] Landes F P, Biroli G, Dauchot O, Liu A J and Reichman D R 2020 Attractive versus truncated repulsive supercooled liquids: the dynamics is encoded in the pair correlation function *Phys. Rev.* E **101** 010602

[52] Banerjee A, Sengupta S, Sastry S and Bhattacharyya S M 2014 Role of structure and entropy in determining differences in dynamics for glass formers with different interaction potentials *Phys. Rev. Lett.* **113** 225701

[53] Nandi M K, Banerjee A, Sengupta S and Bhattacharyya S M 2015 Unraveling the success and failure of mode coupling theory from consideration of entropy *J. Chem. Phys.* **143** 174504

[54] Banerjee A and Wales D J 2020 Fragility and correlated dynamics in supercooled liquids *J. Chem. Phys.* **153** 124501

[55] Rudin C 2019 Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead *Nat. Mach. Intell.* **1** 206–15

[56] Murdoch W J, Singh C, Kumbier K, Abbasi-Asl R and Yu B 2019 Definitions, methods and applications in interpretable machine learning *Proc. Natl Acad. Sci.* **116** 22071–80

[57] Sciortino F and Kob W 2001 Debye-waller factor of liquid silica: theory and simulation *Phys. Rev. Lett.* **86** 648–51

[58] Baity-Jesi M and Reichman D R 2019 On mean-field theories of dynamics in supercooled liquids *J. Chem. Phys.* **151** 084503

[59] Carof A, Vuilleumier R and Rotenberg B 2014 Two algorithms to compute projected correlation functions in molecular dynamics simulations *J. Chem. Phys.* **140** 124103

[60] Jung G, Hanke M and Schmid F 2017 Iterative reconstruction of memory kernels *J. Chem. Theory Comput.* **13** 2481–8

[61] Klippenstein V, Tripathy M, Jung G, Schmid F and van der Vegt N F A 2021 Introducing memory in coarse-grained molecular simulations *J. Phys. Chem.* B **125** 4931–54

[62] Vroylandt H and Monmarché P 2022 Position-dependent memory kernel in generalized Langevin equations: theory and numerical estimation *J. Chem. Phys.* **156** 244105

[63] Bengtzelius U, Gotze W and Sjolander A 1984 Dynamics of supercooled liquids and the glass transition *J. Phys. C: Solid State Phys.* **17** 5915–34

[64] van Megen W and Underwood S M 1993 Glass transition in colloidal hard spheres: mode-coupling theory analysis *Phys. Rev. Lett.* **70** 2766–9

[65] Coslovich D, Ninarello A and Berthier L 2019 A localization transition underlies the mode-coupling crossover of glasses *SciPost Phys.* **7** 77

[66] Carbone M R and Baity-Jesi M 2022 Competition between energy- and entropy-driven activation in glasses *Phys. Rev.* E **106** 024603

[67] Andersen H C 2005 Molecular dynamics studies of heterogeneous dynamics and dynamic crossover in supercooled atomic liquids *Proc. Natl Acad. Sci.* **102** 6686–91

[68] Leutheusser E 1984 Dynamical model of the liquid-glass transition *Phys. Rev.* A **29** 2765–73

[69] Ciarella S, Biezemans R A and Janssen L M C 2019 Understanding, predicting and tuning the fragility of vitrimeric polymers *Proc. Natl Acad. Sci.* **116** 25013–22

[70] Ciarella S 2021 Relaxation pathways for soft materials (available at: https://research.tue.nl/en/publications/relaxation-pathways-for-soft-materials)

[71] Kob W and Andersen H C 1994 Scaling behavior in the β-relaxation regime of a supercooled Lennard–Jones mixture *Phys. Rev. Lett.* **73** 1376–9

[72] Weeks J D, Chandler D and Andersen H C 1971 Role of repulsive forces in determining the equilibrium structure of simple liquids *J. Chem. Phys.* **54** 5237–47

[73] Anderson J A, Lorenz C D and Travesset A 2008 General purpose molecular dynamics simulations fully implemented on graphics processing units *J. Comput. Phys.* **227** 5342–59

[74] Bengio Y 2012 Practical recommendations for gradient-based training of deep architectures (arXiv:1206.5533)

[75] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)

[76] Hansen N and Ostermeier A 2001 Completely derandomized self-adaptation in evolution strategies *Evol. Comput.* **9** 159–95

[77] Hansen N, Akimoto Y and Baudis P 2019 CMA-ES/pycma on github (available at: https://zenodo.org/record/7573532#.ZC5owNJBxkg)