Dynamic heterogeneity at the experimental glass transition predicted by transferable machine learning

Gerhard Jung [1], Giulio Biroli,[2] and Ludovic Berthier [1,3]

[1]*Laboratoire Charles Coulomb (L2C), Université de Montpellier, CNRS, 34095 Montpellier, France*
[2]*Laboratoire de Physique de l'Ecole Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université,
Université de Paris, F-75005 Paris, France*
[3]*Gulliver, UMR CNRS 7083, ESPCI Paris, PSL Research University, 75005 Paris, France*

We develop a machine learning model, which predicts structural relaxation from amorphous supercooled liquid structures. The trained networks are able to predict dynamic heterogeneity across a broad range of temperatures and time scales with excellent accuracy and transferability. We use the network transferability to predict dynamic heterogeneity down to the experimental glass transition temperature T_g, where structural relaxation cannot be analyzed using molecular dynamics simulations. The results indicate that the strength, the geometry, and the characteristic length scale of the dynamic heterogeneity evolve much more slowly near T_g compared to their evolution at higher temperatures. Our results show that machine learning techniques can provide physical insights on the nature of the glass transition that cannot be gained using conventional simulation techniques.

I. INTRODUCTION

Dense liquids display a drastic slowing down of structural relaxation when approaching the experimental glass transition temperature [1,2]. The glass transition is characterized by several important properties, such as a very homogeneous amorphous structure but a strongly heterogeneous relaxation dynamics, leading to the spatial coexistence of frozen and active regions [3]. Understanding the connection between the microstructure and dynamic heterogeneity is an important field of research [4–7].

Over the years, several structural order parameters have been proposed, which show some degree of correlation with the local relaxation dynamics, including density [8], potential energy [9,10], geometry of Voronoi cells [11], soft modes [4], locally-favored structures [5,12–14], and more [6,7,15,16]. Recently, the application of machine learning (ML) techniques to automatically construct suitable structural order parameters has significantly advanced this line of research. The range of methodologies includes unsupervised learning to automatically detect structural heterogeneities [8,17–19] and supervised learning using linear regression [20–22], support vector machines [23–26], multilayer perceptrons (MLP) [27], and graph neural networks (GNN) [28–31]. The performance of these techniques significantly surpasses traditional approaches based on hand-made order parameters, and allows to infer the microscopic structural relaxation from structural properties with high accuracy, including aspects of dynamic heterogeneity [27,29].

These advancements in ML techniques have enabled the attainment of novel physical insights. Leveraging the scalability of trained neural networks with respect to system size facilitated new results on dynamic length scales and the geometry of rearranging domains [27]. Transferability to other state points has been used to analyze structural differences between strong and fragile glass formers [32] or extract system-averaged relaxation dynamics [33]. Trained models were also applied to construct effective glass models [34]. Transferability to lower temperature has also been analyzed in one of the first supervised ML applications [28,29]. It was shown that for GNNs some correlation between structure and dynamics persists when applying the trained networks to different temperatures. For maximum performance, however, the networks need to be trained separately for each state point, thus requiring that training sets already exist at each time and temperature [28,29]. Additionally, a full analysis of transferability is lacking, also with respect to predicting dynamic heterogeneity. The aim of this article is to fill this gap, and present an efficient method to predict physical properties outside of the training regime, possibly including physical regimes that cannot easily be accessed using conventional numerical techniques.

We develop a transferable ML framework, which is able to learn and predict time-dependent dynamical properties when given amorphous structures of deeply supercooled liquids. Different from most previously proposed ML techniques, the network is trained using data extracted from very different temperatures and time scales, and is thus able to maintain both an excellent performance over the whole range of data provided. We show that the network can be transferred to predict relaxation beyond the range of temperatures provided during training. As an important application, we create equilibrium structures down to the experimental glass transition temperature T_g, using the swap Monte Carlo algorithm [35,36] and apply the transferable network to investigate dynamic susceptibilities and lengths scales at these very low temperatures, which correspond to time scales that are not accessible by molecular dynamics (MD) simulations.
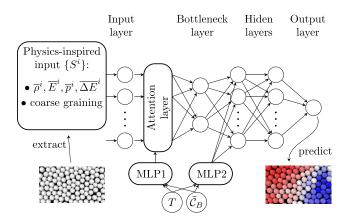
FIG. 1. Geometry of the transferable machine learning model, which we name "tGlassMLP". The various layers and their respective role are described in Sec. II.

The paper is organized as follows. First, the transferable ML methodology is described in Sec. II. The performance of the trained network is analyzed in Sec. III, both for state points within the range of training data and beyond to study transferability. In Sec. IV we use the transferability to obtain results for structural relaxation at the experimental glass transition temperature. In Sec. V, we analyze in detail the dynamic correlation lengths of our system. To better understand the trained network, we study some of its intrinsic properties, such as inherent length scales and properties of the bottleneck layer in Sec. VI. We discuss the results and conclude in Sec. VII.

II. MACHINE LEARNING METHODOLOGY

The transferable ML technique presented in this paper is a generalization of the GlassMLP network introduced in Ref. [27], which we modify and improve to enable transferability. We will refer to this model as "transferable GlassMLP" and use the acronym "tGlassMLP" for the model. The geometry of tGlassMLP is sketched in Fig. 1. It is comprised of an input layer, an attention layer, and a dense multilayer perceptron (MLP) network including a bottleneck layer. We now describe these individual parts, define the glass-forming system under study and the dynamical observables used to train the network.

A. Physical system and dynamical observables

We study the same system and dynamical observables used in Ref. [27], but we briefly recapitulate the most important details and definitions to make the paper self-contained.

We investigate a two-dimensional ternary glass former, which generalises the Kob-Andersen binary mixture (KA2D) [37]. To enable fast equilibration using the swap Monte Carlo algorithm [35,36,38] we include a third particle type with an intermediate size [39]. We use reduced units, which are defined in terms of length $\sigma = 1$ (corresponding to the size of the large particles), mass $m = 1$ (mass of the particles), and energy $\epsilon = 1$ (Lennard-Jones energy scale of the interactions between large particles). In these units, the standard system size is $L_S = 32.896$, with periodic boundary conditions, in which we create amorphous packings of $N = 1290$ particles

($N_1 = 600, N_2 = 330, N_3 = 360$), if not otherwise stated. Using swap Monte Carlo, we create equilibrium configurations in a temperature range between $T = 0.4$, which is slightly below the onset temperature ($T_{\rm on} \approx 0.5$), and the estimated experimental glass transition temperature $T_g = 0.15$ (see SM in Ref. [27] for details). Equilibrium averages over these configurations are denoted as $\langle \ldots \rangle$. No sign of crystallization was observed even at T_g. Starting from the equilibrated configurations we perform molecular dynamics (MD) simulations for five different temperatures, $T = 0.4$, $T = 0.3$, $T = 0.25$, $T = 0.23$, and $T = 0.21$, to investigate the physical relaxation dynamics. We analyze the isoconfigurational average [4,40], which means that we perform $N_R = 20$ different simulations starting from each structure (denoted as replicas), which are created by randomly drawing initial velocities from the Maxwell distribution. The isoconfigurational average $\langle \ldots \rangle_{\rm iso}$ for a given initial structure denotes the average over these N_R replicas.

The labels used to train tGlassMLP are extracted from the MD trajectories using the bond-breaking correlation function (BB). The isoconfigurational average of BB, which we refer to as "propensity" in the following [4,40], is defined as $\mathcal{C}_B^i(t) = \langle n_t^i/n_0^i \rangle_{\rm iso}$, where n_0^i denotes the initial number of neighbors of particle i, and n_t^i the number of these neighbors, which remain neighbors of i after a time t [41]. Therefore, $\mathcal{C}_B^i(t) \in [0, 1]$, where $\mathcal{C}_B^i(t) = 1$ denotes arrested particles, which did not loose a single neighbor in any of the replicas (visualized as blue in all snapshots) and $\mathcal{C}_B^i(t) \ll 1$ denotes very active particles (visualized as red). From the averaged propensity $\bar{\mathcal{C}}_B(t) = \frac{1}{N_1} \sum_{i=1}^{N_1} \mathcal{C}_B^i(t)$, we extract a bond-breaking structural relaxation time $\tau_\alpha^{\rm BB}$, which is defined as $\langle \bar{\mathcal{C}}_B(t = \tau_\alpha^{\rm BB}) \rangle = 0.5$. More details are documented in the SM of Ref. [27]. We mostly report results for particles of type 1 but verified that all findings are independent of the particle type. We calculate BB from MD simulations for each temperature introduced above at various times, yielding many sets of labels at different state points for the supervised learning procedure of tGlassMLP.

To describe the state point of each set of labels, we use the temperature T and the averaged propensity $\langle \bar{\mathcal{C}}_B(t) \rangle$. In particular, using $\langle \bar{\mathcal{C}}_B(t) \rangle$ instead of the time t itself is a crucial choice due to a near perfect time-temperature superposition, i.e., perfect collapse of the function $\langle \bar{\mathcal{C}}_B(t/\tau_\alpha^{\rm BB}) \rangle$ measured at different temperatures. Interchanging time for the value of $\langle \bar{\mathcal{C}}_B(t) \rangle$ simplifies the training procedure significantly since $\langle \bar{\mathcal{C}}_B(t) \rangle \in [0, 1]$, while t grows exponentially with decreasing temperature. To translate $\langle \bar{\mathcal{C}}_B(t) \rangle$ back into a time, we assume that time-temperature superposition continues to hold at any temperature, combined with the extrapolated value for $\tau_\alpha^{\rm BB}$. More details of this procedure are provided in Appendix A. While this assumption might affect slightly the time dependence of the ML predictions presented in Secs. IV–VI, they do not change the quality of the predictions themselves. Thus, the above encoding of time does not affect our analysis of transferability.

B. Physics-inspired structural input

The structural input is the same as for GlassMLP [27]. In particular, we use M_S physics-inspired and coarse-grained input features for the description of the local structure of

each particle i. The descriptors are based on $K = 4$ different observables:

(1) The coarse-grained local density:

$$\overline{\rho}^i_{L,\beta} = \sum_{j \in N^i_\beta} e^{-R_{ij}/L}, \qquad (1)$$

which sums over the N^i_β particles of type β within distance $R_{ij} = |\boldsymbol{R}_i - \boldsymbol{R}_j| < 20$ of particle i.

(2) The coarse-grained potential energy:

$$\overline{E}^i_{L,\beta} = \sum_{j \in N^i_\beta} E^j e^{-R_{ij}/L}/\overline{\rho}^i_{L,\beta}, \qquad (2)$$

extracted from the pair interaction potential $E^i = \sum_{j \neq i} V(R_{ij})/2$.

(3) The coarse-grained Voronoi perimeter:

$$\overline{p}^i_{L,\beta} = \sum_{j \in N^i_\beta} p^j e^{-R_{ij}/L}/\overline{\rho}^i_{L,\beta}, \qquad (3)$$

based on the perimeter p^i of the Voronoi cell around particle i, extracted using the software Voro++ [42].

(4) The local variance of potential energy:

$$\overline{\Delta E}^i_{L,\beta} = \sum_{j \in N^i_\beta} (E^j - \overline{E}^i_{L,\beta})^2 e^{-R_{ij}/L}/\overline{\rho}^i_{L,\beta}. \qquad (4)$$

Particle positions are evaluated in inherent structures $\boldsymbol{R}_i$. Slightly differently from Ref. [27], we use only $M_{CG} = 11$ different values of the coarse-graining lengths L, which are nonuniformly distributed, $L = \{0.0, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 7.0, 9.0\}$. The four descriptors are separately coarse-grained by iterating over each of the $M_{type} = 3$ particle types. We additionally calculate the coarse-grained average by running over all particles independently of their type.

Overall, we start with a set of $M_S = KM_{CG}(M_{type} + 1) = 176$ descriptors. To enable more efficient training, each descriptor is shifted and rescaled to have zero mean and unit variance.

C. Attention layer

The first real difference between tGlassMLP and its GlassMLP ancestor is the introduction of an attention layer between the input and bottleneck layers (see Fig. 1). The concept of attention and transformers have increased significantly the performance of many ML models in computer science [43], and it has already been used in glass physics for the development of an improved GNN [31]. The purpose of the attention layer in tGlassMLP is to learn a state-dependent weight w_j, which is assigned to each structural descriptor j. In this way, the dependence of the dynamical descriptor on the considered state point is efficiently encoded in the network.

If we denote the M_S values of the physics-inspired descriptors by $\{S^1, ..., S^{M_S}\}$, then the attention layer can be written as

$$A^k = S^k f_1^k + \text{MLP1}^1_{\text{out}} f_2^k + \text{MLP1}^2_{\text{out}} f_3^k + f_4^k, \qquad (5)$$

$$w_k = \text{softmax}(\{A^1, \ldots, A^{M_S}\})_k, \qquad (6)$$

$$S^k_{\text{out}} = S^k w_k. \qquad (7)$$

Here, f_n^k denote the learnable parameters ($4 \times M_S = 688$ in total) and $\text{MLP1}^n_{\text{out}}$ is the two-dimensional output of a small MLP (denoted as MLP1). The softmax function ensures that the weights are normalized. The output of the attention layer still has dimension M_S, where each input descriptor, S^k is multiplied by its specific weight w_k. The attention layer is able to reweight the input before encoding it in the bottleneck layer. It can thus learn, for example, that the relative importance of different structural indicators depends on time and temperature. We will explicitly analyze these weights w_k, obtained after training tGlassMLP, in Sec. VI to extract meaningful physical information from interpreting the network itself.

D. Dense MLP with bottleneck

Similar to GlassMLP, after the attention layer, the high-dimensional input is encoded into a two-dimensional bottleneck layer, to avoid having a huge amount of free parameters in the subsequent hidden layers, which would lead to overfitting. The optimal choice of a two-dimensional geometry has been found empirically by testing different bottleneck sizes. As for GlassMLP [27] we have found that a one-dimensional description is not sufficient to maintain the same level of predictability, while three and more dimensions included additional fitting parameters without leading to significant improvement. We visualize and interpret the bottleneck layer of trained tGlassMLP networks in Sec. VI.

After the bottleneck layer, the state point $(T, \langle \overline{\mathcal{C}}_B(t) \rangle)$ is again explicitly inserted into the network using a second small MLP (denoted as MLP2). The bottleneck layer concatenated with the output of MLP2 will then be further processed in two hidden layers to yield the final output (see Fig. 1).

E. Two-step training of tGlassMLP

The total number of free parameters of tGlassMLP is slightly above 1000. This number is several orders of magnitude less than for the GNNs proposed in Refs. [28–31]. Importantly also, this number is not much larger than in the original version of GlassMLP where about 650 fitting parameters were used [27]. To be as efficient as GlassMLP, which was separately trained at each state point, the tGlassMLP network is therefore inherently forced to learn universal aspects in the structural relaxation across time scales and temperatures. This appears instrumental to construct a model with good transferability.

To train the network, we use a supervised ML procedure, in which the output of tGlassMLP $\mathcal{X}^i_{\text{MLP}}$ is rated by a differentiable loss function, which is the same as used in Ref. [27]. It includes the mean absolute error, but also additional terms, which penalize deviations between the predicted and the true variance, as well as spatial correlations of the propensities (see SM of Ref. [27]). We use $N_S = 300$ initial structures, which are equally divided into a training and a test set. For the training we apply stochastic gradient descent with an Adam optimizer [44].

The training of tGlassMLP is performed in two steps. First, we train in an "equal-time" mode, meaning that separate networks are trained for given values of $\langle \overline{\mathcal{C}}_B(t) \rangle$ (i.e., at equal times relative to the structural relaxation time), but different

temperatures $T \geqslant T_{\min}$. We found that these individual networks transfer better to lower temperatures than the ones who were directly trained on all state points.

These individual networks are then applied to lower temperatures $T < T_{\min}$ and the average of the predicted propensity $\overline{\mathcal{X}_{\mathrm{MLP}}} = \frac{1}{N_1} \sum_i^{N_1} \mathcal{X}_{\mathrm{MLP}}^i$ is calculated. We have found that depending on the initial condition of the training, not all resulting networks are equally efficient. We keep all networks, which fulfill $|\overline{\mathcal{X}_{\mathrm{MLP}}} - \langle \bar{\mathcal{C}}_B(t) \rangle| < 10^{-4}$ for further processing. This is a self-consistency test, as the average predicted propensity is an input data. The networks, which satisfy the above criterion are used to predict propensities for the low temperature configurations $T < T_{\min}$ at the values of $\langle \bar{\mathcal{C}}_B(t) \rangle$ for which MD results are not available [typically low values of $\langle \bar{\mathcal{C}}_B(t) \rangle$]. Similarly to knowledge distillation in machine learning, we include them into the training procedure of the full and final tGlassMLP model. See Table I in the Appendix B for a more detailed description.

In the second training step, a single tGlassMLP network is trained using data from all times and temperatures, including the extrapolated data at $T < T_{\min}$ produced in the first step. As before, we only retain networks such that $|\overline{\mathcal{X}_{\mathrm{MLP}}} - \langle \bar{\mathcal{C}}_B(t) \rangle| < 10^{-4}$ for predicted propensities at $T < T_{\min}$. Out of 16 networks trained initially in this way in the second step, four networks were selected for the predictions shown in the remainder of this paper. Error bars on dynamic quantities calculated from the predictions of tGlassMLP correspond to the variance between the predictions made by these individual networks. Large error bars, in particular for extrapolated results, are therefore an indication of strong overfitting.

Once the network is trained, predictions are extracted on the test set for all configurations at the different target temperatures T. For this task, the physics-inspired structural input is calculated for each configuration and fed into the network along with the temperature T and a series of $\langle \bar{\mathcal{C}}_B(t) \rangle$. Since this input is continuous, predictions of tGlassMLP will be represented as continuous lines in the following.

Additional details on hyperparameters, values for $\langle \bar{\mathcal{C}}_B(t) \rangle$ and temperatures used for training are presented in Appendix B. The training of tGlassMLP was performed on a Laptop GPU (NVIDIA T600 Laptop) and with a total computational cost of about one day. This includes the training of all 16 individual networks per time scale in the first training step, and of 16 networks using the full data for all times and temperatures in the second step.

In the following we present our results, which are organized in four parts. In the first part we validate the performance and the transferability of tGlassMLP using results known from MD simulations. In the second part, we use the trained models to predict structural relaxation at the experimental glass transition temperature, which is not accessible by computer simulations. The third and fourth parts include a detailed analysis of dynamic correlation lengths and of the properties of the trained tGlassMLP network itself.

III. VALIDATION OF tGLASSMLP

We follow common practice [28] and investigate the performance of the trained model using the Pearson correlation
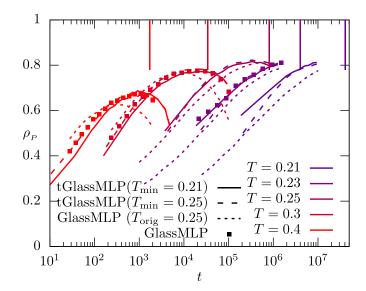

FIG. 2. Pearson correlation coefficient of several trained ML models with MD simulations. Squares correspond to results shown in Ref. [27] for GlassMLP trained at each state point. Dotted lines correspond to a GlassMLP model trained at $t = \tau_\alpha$ and $T_{\mathrm{orig}} = 0.25$ and transferred to different times and temperatures. Full and dashed lines describe results for two transferable tGlassMLPs trained until minimum temperatures $T_{\min} = 0.21$ and $T_{\min} = 0.25$. Vertical lines indicate the structural relaxation times $\tau_\alpha^{\mathrm{BB}}$ at each temperature.

coefficient,

$$\rho_P = \mathrm{cov}\big(\mathcal{C}_B^i, \mathcal{X}_{\mathrm{MLP}}^i\big) / \sqrt{\mathrm{var}\big(\mathcal{C}_B^i\big)\mathrm{var}\big(\mathcal{X}_{\mathrm{MLP}}^i\big)}, \qquad (8)$$

which quantifies the correlation between the predicted propensities and the ground truth extracted from the MD simulations. The Pearson correlation varies between $\rho_P = 1$ (perfect correlation) and $\rho_P = -1$ (perfect anticorrelation), while no correlation yields $\rho_P = 0$.

We find that the maximum of the Pearson correlation reaches values of roughly $\rho_P^{\max} \approx 0.8$ for the lowest temperatures, thus indicating very strong correlations (see Fig. 2). Most importantly, we observe that the performance of the trained tGlassMLP networks (full lines) is as good as the results reported in Ref. [27] in which the networks were individually trained on each state point (squares). This is a significant result considering that the tGlassMLP network only uses twice the number of fitting parameters to describe more than 50 different state points. This shows that the transferable network indeed discovers common features in the description of structural relaxation from the microscopic structure at different times and temperatures. The performance of the model is similar to the one of other recently proposed techniques such as CAGE [22], BOTAN [30], and the SE(3) extension to GNNs [29]. For comparison of the Pearson correlation to other ML techniques we refer to a recent roadmap [45], which presents benchmarks based on similar data as used for this paper (KA2D $T = 0.3$).

We have also trained tGlassMLP using a smaller dataset down to $T_{\min} = 0.25$ (dashed line in Fig. 2). The model very favorably transfers to lower temperatures and can predict structural relaxation on time scales that are orders of
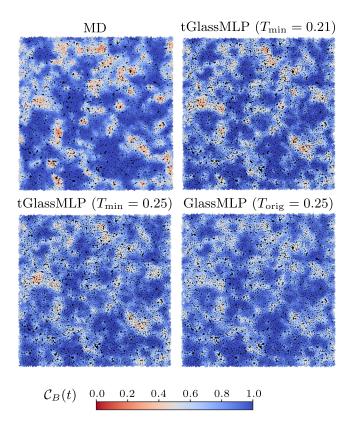
FIG. 3. Snapshots comparing the spatial variations of the structural relaxation at $T = 0.21$ and $\langle \bar{C}_B \rangle = 0.8$ obtained in MD simulations, to three different models: tGlassMLP networks trained using $T_{\min} = 0.21$ and $T_{\min} = 0.25$, and GlassMLP trained at $T_{\text{orig}} = 0.25$ and $t = \tau_\alpha^{\text{BB}}$.
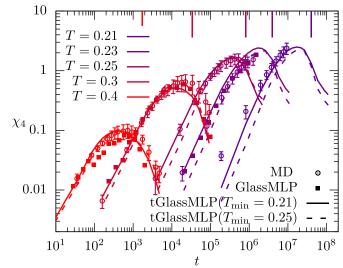

FIG. 4. Dynamic four-point susceptibility $\chi_4(t)$. Points are MD results, squares are results shown in Ref. [27] for GlassMLP trained at each state point. Full and dashed lines correspond to transferable tGlassMLPs trained until minimum temperatures $T_{\min} = 0.21$ and $T_{\min} = 0.25$. Vertical lines indicate τ_α^{BB} at each temperature. Here and in the following, error bars of MD results correspond to the variance between three independent datasets, each averaged over 100 independent structures.

magnitude longer than in the training dataset with nearly as much accuracy as the directly trained model. This demonstrates excellent transferability of tGlassMLP in the regime where this can be tested quantitatively.

In addition, tGlassMLP outperforms the directly transferred GlassMLP model, which was trained at the state point $t = \tau_\alpha$ and $T_{\text{orig}} = 0.25$ (dotted lines). This conclusion similarly holds when comparing the results to other ML techniques such as GNNs [28,29]. In particular, the equivariant network proposed in Ref. [29] shows very comparable transferability as the original GlassMLP network [27], but cannot match the improved transferability performance of tGlassMLP.

To visualize the correlation between the propensities obtained from MD simulations and the different ML models we show snapshots for a large configuration with $N = 25800$ at $T = 0.21$ in Fig. 3. The tGlassMLP models are very accurate in predicting the location of both strongly rearranging regions and frozen regions in which no rearrangements take place. Stronger differences with the MD result are observed when using the GlassMLP network trained at $T_{\text{orig}} = 0.25$ and transferred to $T = 0.21$. While also for this model, the propensities are correlated with the MD results, dynamic heterogeneities are less pronounced and the contrast between active and frozen regions is not captured properly. This analysis shows that the quality of a given ML model is not uniquely quantified by a strong Pearson correlation, but that accurate modeling of additional glassy properties, for example, also requires

correct representation of dynamic heterogeneities, which is a nontrivial task.

To analyze spatial correlations of the propensity quantitatively, we compute the four-point susceptibility $\chi_4(t)$,

$$\chi_4(t) = N_1 \left(\langle \bar{C}_B^2(t) \rangle - \langle \bar{C}_B(t) \rangle^2 \right), \tag{9}$$

which was used extensively to characterize dynamic heterogeneity in supercooled liquids [3]. As expected after inspection of the snapshots in Fig. 3, we find very good agreement between the susceptibilities extracted from the MD results and the trained ML model (see Fig. 4). The results are as good as those reported in Ref. [27] for GlassMLP models trained at each state point. In fact, for $T = 0.4$, the transferable tGlassMLP model even surpasses the performance of GlassMLP, despite being much more broadly applicable to dynamic susceptibilities, which can differ by more than two orders of magnitude in amplitude.

This conclusion holds for the two models trained using different values of $T_{\min}$. Therefore, the tGlassMLP model is able to predict a realistic increase in dynamic heterogeneity with decreasing the temperature even when it is extrapolated beyond the range used in the training set. This excellent transferability sets it apart from techniques, which are trained on a single state point and are therefore unable to capture variations in the overall dynamic heterogeneity. The tGlassMLP models also predict a realistic decay of the susceptibility for times longer than the structural relaxation time even though for $T < 0.25$ no training data for times $t > 0.3\tau_\alpha^{\text{BB}}$ were used during training. The models are therefore not only transferable in temperature, but also in time.

The above analysis clearly highlights that the tGlassMLP network is able to extrapolate its predictions outside of the

range covered in the training set. It is therefore tempting to train tGlassMLP with as much data as is currently possible using MD simulations, and then use the network to predict features of structural relaxation at times and temperatures where MD simulations can no longer be performed. This is investigated in the next section.

IV. PREDICTING THE DYNAMICS AT THE EXPERIMENTAL GLASS TRANSITION TEMPERATURE

The glass transition temperature T_g is conventionally defined as the temperature at which structural relaxation occurs roughly 10^{12} times slower than in the simple liquid. Simulating the relaxation of supercooled liquids at T_g would require integrating about 10^{14} time steps. The computational cost of the data provided in this section would thus roughly amount to 10^{11} CPU hours. Equivalently, this would require about 100 days of simulation by completely exhausting the Top 500 supercomputers in the world.

Our approach is obviously more parsimonious and combines two algorithms: (i) the swap Monte Carlo algorithm allows us to efficiently create independent equilibrium configurations of our glass model down to T_g, and (ii) the transferable tGlassMLP model, which can predict dynamic propensities for each particle from the equilibrium configurations obtained in (i). This unique combination enables us to predict and analyze the strength, geometry, and length scale of dynamic heterogeneity at unprecedentedly low temperatures and large times.

In Fig. 5 we show snapshots describing the predicted time evolution of the structural relaxation in a given sample at $T = T_g$ with $N = 82560$ using the tGlassMPL model. The same process is displayed as a movie in the SM, which better highlights how relaxation sets in at specific localized regions within the amorphous structure before slowly spreading over the entire system. This qualitative description is similar to results produced by MD simulations performed for a lesser degree of supercooling.

Since these predictions are made for temperatures where direct MD simulations can no longer be performed, there is no way to directly test the quality of these predictions. The plausibility of the result is guaranteed by the excellent transferability demonstrated in the previous section for a higher temperature regime and the physically consistent behavior that is predicted.

We quantify the dynamic heterogeneities visible in these snapshots using the four-point susceptibility $\chi_4(t)$, defined in Eq. (9). The results are presented in Fig. 6. The time dependence is extracted by predicting the value of C_B^i for each particle for a given value of the average quantity $\langle \bar{\mathcal{C}}_B \rangle$, and converting $\langle \bar{\mathcal{C}}_B \rangle$ into a time using time-temperature superposition, as discussed above in Sec. II A. The predicted $\chi_4(t)$ functions continue to have the nonmonotonic time dependence they have at higher temperatures, and we note that the amplitude of the maximum of χ_4 grows very modestly for temperatures $T < 0.21$ (see Fig. 6). It should be emphasized that this is not a trivial result. Indeed in Sec. III we have observed that tGlassMLP with $T_{min} = 0.25$ has correctly predicted a slowly-increasing χ_4 when transferred beyond the temperature range it has been trained on.
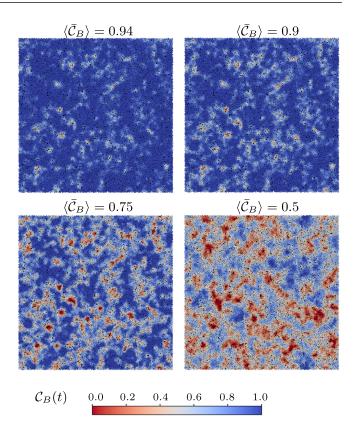

FIG. 5. Snapshots visualizing the time evolution of the structural relaxation at $T_g = 0.15$, with the corresponding value of the average $\langle \bar{\mathcal{C}}_B \rangle$ indicated, corresponding to estimated times $t = 10^{11}, 3 \times 10^{11}, 10^{12}, 3 \times 10^{12}$. The corresponding movie can be found within the Supplemental Material [46].

From the time-dependent susceptibility we can extract for each temperature the maximum in time $\max_t \chi_4(t) = \chi_4^{max}$ and the corresponding time scale $\chi_4(\tau_{\chi_4^{max}}) = \chi_4^{max}$. In practice, we extract the maximum by fitting a log-normal
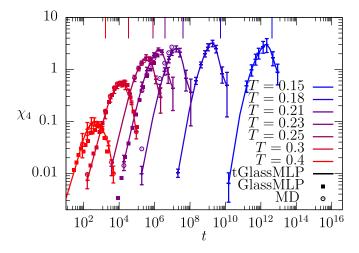

FIG. 6. Dynamic four-point susceptibility $\chi_4(t)$ for tGlassMLP (using $T_{min} = 0.21$). Square data points correspond to results shown in Ref. [27] and points to MD results. Vertical lines indicate τ_α^{BB} at each temperature. Error bars of tGlassMLP correspond to the variance between individual networks, trained using the same dataset but with different initial weights, as described in Sec. II E.
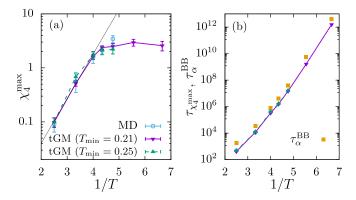
FIG. 7. (a) Maximal values $\chi_4^{max}(T)$ as shown in Figs. 4 and 6 for each temperature extracted from tGlassMLP (tGM), using $T_{min} = 0.21$ and $T_{min} = 0.25$, as well as MD simulations. The black line corresponds to an exponential increase and serves as guide for the eye. (b) Time scale of the maximal values $\tau_{\chi_4^{max}}(T)$ for the same data, compared to the structural relaxation time τ_α^{BB}.

distribution to $\chi_4(t)$ around the maximum, which describes the temporal evolution very well. For the two temperatures $T = 0.23$ and $T = 0.21$ our MD data does not reach a maximum, we therefore fix the time scale $\tau_{\chi_4^{max}} = \tau_\alpha^{BB}/2$ to extract an estimate for χ_4^{max}.

Figure 7(a) confirms our previous discussion in highlighting the crossover predicted by tGlassMLP in $\chi_4^{max}(T)$ at a temperature around $T = 0.25$, which is slightly below the mode-coupling temperature. While the MD data also show a crossover within the accuracy of our measurement it is not obvious how strongly $\chi_4^{max}(T)$ will continue to grow beyond $T < 0.25$. Additional simulations, also for longer times, to reduce the statistical errors would be necessary to obtain a clearer picture. In contrast to the predictions by tGlassMLP such measurements would, however, be computationally extremely expensive. The situation is much clearer for the timescale $\tau_{\chi_4^{max}}$, which very consistently follows $\tau_{\chi_4^{max}} = \tau_\alpha^{BB}/2$ for all methods investigated [see Fig. 7(b)]. This observation also justifies the above choice for fixing the time scale of the maximum for the low temperature MD data.

V. DYNAMIC LENGTH SCALES

A. Four-point dynamic length scale

The dynamic heterogeneities visualized in Fig. 5 and quantified via $\chi_4(t)$ in Fig. 6 can also be characterized by a dynamic correlation length, which describes spatial correlations of the dynamic propensity. Calculating static and dynamic length scales, in particular at very low temperatures, is of high importance in glass physics, since the emergence and growth of dynamic correlation length scales is a key element of various different theories of the glass transition [47–50].

Here, we first adopt the same methodology as used in Ref. [27] to extract a dynamic length scale from the four-point dynamic structure factor [48],

$$S_4(q, t) = N_1^{-1} \langle W(\boldsymbol{q}, t) W(-\boldsymbol{q}, t) \rangle,$$
$$W(\boldsymbol{q}, t) = \sum_{i \in N_1} \left(\mathcal{C}_B^i(t) - \langle \bar{\mathcal{C}}_B(t) \rangle \right) \exp[i\boldsymbol{q} \cdot \boldsymbol{R}_i(0)]. \quad (10)$$
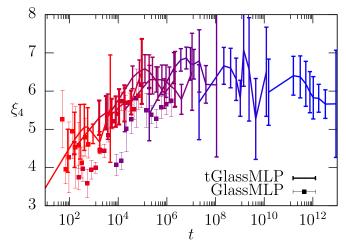

FIG. 8. Four-point dynamic length scale $\xi_4(t)$ extracted from $S_4(q, t)$. Results are shown for the tGlassMLP model with $T_{min} = 0.21$. We also show the length scale extracted in Ref. [27] using GlassMLP. Colors are the same as in Fig. 6.

The dynamic structure factor $S_4(q, t)$ characterizes the geometry and spatial extent of the regions defined by correlated fluctuations of $\mathcal{C}_B^i(t)$. In the limit $q \to 0$, it has been predicted to decay quadratically with $1/q$, according to the Ornstein-Zernicke form [48],

$$S_4(q, t) \approx \frac{\tilde{\chi}_4}{1 + (q\xi_4(t))^2}. \quad (11)$$

Here, $\tilde{\chi}_4$ is connected (but not equal) to the dynamical susceptibility χ_4 [48] defined above, and $\xi_4(t)$ denotes the time-dependent dynamic length scale. This correlation length can be extracted by fitting the measured dynamic structure factor for small q. Recently, we have applied this method to GlassMLP networks by utilizing the scalability of the network in system size [27]. This last step was necessary, since an accurate extraction of ξ_4 requires very large system sizes with linear size $L_S \gg \xi_4$ [48,51–53]. Here, we apply the same methodology as in Ref. [27], using the same fitting parameters, to extract the dynamical length scale ξ_4 from systems with $N = 82560$ particles, but we now use tGlassMLP models. The idea is that we can efficiently generate equilibrium configurations for such large systems using swap Monte Carlo [35] and use the tGlassMLP model to predict the emergent dynamic propensity. Consistent with Ref. [54] we have shown in the SI of Ref. [27] that this procedure is not affected by strong finite size effects and allows us to extract dynamical information from large systems.

The results for ξ_4 are shown in Fig. 8. At any given temperature, the length scale increases with time up to a maximum value, which is located roughly at the structural relaxation time τ_α^{BB}. This maximum increases steeply with decreasing the temperature up to $T = 0.25$, below the growth is much less pronounced to reach maximal values of the order $\xi_4 \approx 6 - 7$ at most. This result is consistent with the findings in Ref. [27] for $T \leqslant 0.23$, and is also consistent with the slow increase of the dynamic susceptibility χ_4 in Fig. 6.

Similarly to Ref. [27] we have also extracted a higher-order contribution to the dynamic structure factor, as shown

in Appendix C. Its evolution with time and temperature again mirrors the behavior of χ_4 and ξ_4.

B. Average chord length

To enable a more detailed comparison of these results to MD simulations we propose an alternative definition for a dynamic length scale, which is connected to the average chord length recently defined in Ref. [50]. One advantage of the chord length is that it does not require very large system sizes, in contrast to the extraction of ξ_4, which does [48,51–53]. To calculate the average chord length, particles are mapped onto a discrete lattice, and the propensity a of each lattice site is defined as the average propensity of the particles within the lattice site. If the propensity of a site is $a < 0.5$, it is denoted as mobile, and immobile otherwise. After discretization and thresholding, chords are defined as series of adjacent mobile sites along all rows and columns of the lattice. The linear size of a chord is ξ_{chord}, and the average over all chords provides the average chord length $\langle \xi_{\text{chord}} \rangle$. The average chord length $\langle \xi_{\text{chord}} \rangle$ is a good measure for the average cluster size, but does not require a cluster analysis. See Sec. VI D and Fig. 18 in Ref. [50] for further explanations and visualization.

We adapt the definition of the lattice propensity a to make the resulting chord length more comparable to ξ_4, in particular for times of the order of τ_α^{BB}. For each time and temperature, we calculate the median of the bond-breaking propensities for all particles. If a particle propensity is below this median it is defined as active ($a^i = 0$), and passive otherwise ($a^i = 1$). Afterwards, we follow the same procedure as described above by mapping the newly defined dynamical quantity a^i onto a lattice, thus redefining the lattice propensity a, and calculating the chord length ξ_{chord}. This leads to a different time evolution compared to the definition in Ref. [50], which features a monotonic growth of the average chord length with time as relaxed regions gradually fill the entire system. In the present version, only the clustering of the mobile particles contributes to the calculation of the chord length.

We show the results for the time and temperature evolution of $\langle \xi_{\text{chord}} \rangle$ in Fig. 9. As anticipated, the average chord length behaves similarly to ξ_4 (compare with Fig. 8). Both quantities display a nonmonotonic dependence in time, although the nonmonotonicity is less pronounced for $\langle \xi_{\text{chord}} \rangle$. Most importantly, both length scales display the same temperature dependence when evaluated in the order of the structural relaxation time τ_α^{BB}. This result therefore establishes $\langle \xi_{\text{chord}} \rangle$ as an interesting measure to extract dynamic length scales, also from smaller systems and for times $t < \tau_\alpha^{\text{BB}}$. Consequently, we can also calculate $\langle \xi_{\text{chord}} \rangle$ from the MD simulations performed to train tGlassMLP. The MD results are in good agreement with the predictions of the ML model (see Fig. 9). The MD length scales are systematically slightly larger than the ones extracted using tGlassMLP, which is likely caused by small inaccuracies in the ML predictions. Importantly, however, both methods clearly display a strong crossover at around $T = 0.25$. The MD results therefore draw the same picture of a dynamic correlation length, which does not grow significantly beyond $\langle \xi_{\text{chord}} \rangle > 14$.
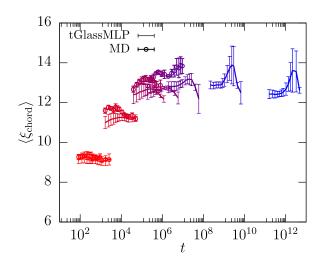

FIG. 9. Average chord length $\langle \xi_{\text{chord}} \rangle$, which can be used to measure dynamic length scales. Colors are the same as in Fig. 6. For each temperature we show times in the range $0.95 > \langle \bar{\mathcal{C}}_B(t) \rangle > 0.3$.

VI. EXTRACTING PHYSICAL INFORMATION FROM THE NETWORK

A natural question when applying machine learning to science is whether one can interpret what the neural network learns in order to achieve its task, and whether it is possible to extract some physical information from the trained network. In the following we will focus on this question for tGlassMLP.

A. Attention layer

We presented in Sec. II the structure of the tGlassMLP network, where we emphasized the addition of an attention layer, which enables tGlassMLP to adapt the weight of different structural descriptors depending on the temperature. We now study these weights explicitly, and relate the observed evolution to the dynamic length scales discussed in Sec. V.

After training, each of the physics-inspired structural input descriptors is assigned a statistical weight, which depends on time and temperature. To relate these weights to a length scale, we first calculate the particle-averaged weight of each descriptor, and average this result over all descriptors, which are coarse-grained over the same length scale L. The resulting normalized averaged weights for each L value are then used to construct the function $P_{\text{int}}(L)$. This function is estimated after training the network, and depends on temperature T.

The temperature evolution of the distribution $P_{\text{int}}(L)$ is shown in Fig. 10. One can clearly observe how the network puts increasing weight on larger length scales when the temperature is decreased. This evolution demonstrates that descriptors, which are coarse grained over larger length scales have an increased weight at lower temperatures. This implies that the network learns the existence of a length scale growing at low temperature and thus determining the range over which structural heterogeneities matter for dynamical behaviors.

We have studied this length scale $\xi_{\text{int}}(T) = \int dL P_{\text{int}}(L) L$ [recall that $P_{\text{int}}(L)$ is normalized]. The temperature evolution of $\xi_{\text{int}}(T)$ compares well to the other dynamic length scales, as shown below in Fig. 13 and discussed further in Sec. VII. This
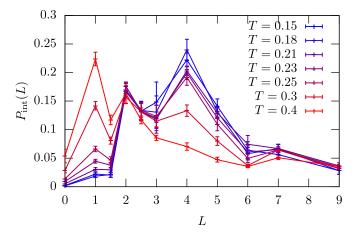
FIG. 10. Normalized distribution of statistical weights $P_{\text{int}}(L)$, determined from the average weights of the attention layer of the tGlassMLP model with $T_{\text{min}} = 0.21$ for each coarse-graining length L. Results are shown for different temperatures T and a single time corresponding to $\langle \bar{\mathcal{C}}_B \rangle = 0.8$. These distributions can be used to define an intrinsic, temperature-dependent dynamic length scale $\xi_{\text{int}}(T) = \int dL P_{\text{int}}(L) L$.

result confirms, *a posteriori*, the utility of the attention layer in tGlassMLP to represent different temperatures with different dynamical correlation lengths. Additionally, it gives an interesting interpretation from the perspective of glass physics to the weights learned by the model.

B. Bottleneck layer

Following the attention layer, the structural input is encoded into a two-dimensional bottleneck layer. This dimensional reduction implies that the dynamics mainly depends on two variables. However, those two variables are obtained by the nontrivial aggregation of several physical local features. Hence, it is unclear how to provide a simple interpretation of those two variables. It is even possible that such an interpretation does not exist.

Still, we can provide some direct evidence of the relationship between these two variables and the dynamical behavior. The output of the bottleneck layer for $T = 0.25$ and different time scales $\langle \bar{\mathcal{C}}_B(t) \rangle$ is shown in Fig. 11. Each point in the figure represents a single particle, its coordinates correspond to the values of the bottleneck layer while the color is used to code for the value of the particle propensity extracted from MD simulations. In this representation, we can observe a clear separation of structure into active and passive particles. All the intricacies of amorphous structure are therefore hidden in the complex encoding from the structural descriptors to the bottleneck layer. Interestingly, in this description structural relaxation over time is "simple". Indeed, the figure clearly demonstrates how the red color, corresponding to the most mobile particles, gradually spreads from the bottom left corner towards the top right. Thus, at early times, only those particles with small values in both bottleneck nodes are likely to rearrange. At longer times, more particles become mobile even at larger values of the bottleneck nodes, until eventually all particles rearrange at very long times.
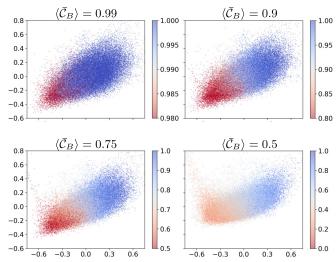

FIG. 11. Visualization of the amorphous structure, encoded by the two-dimensional bottleneck layer, at $T = 0.25$ for different times. Each point represents a single particle ($N = 90000$ particles shown in total) at coordinates given by the value of the bottleneck layer. Each point is colored according to the value of the particle propensity $\mathcal{C}_B^i(t)$, as indicated by the color bar.

We also study the temperature dependence of the bottleneck layer for a fixed timescale relative to τ_α^{BB} in Fig. 12. The aggregate output of the bottleneck layer is slightly deformed by the change in temperature, which arises from the temperature-dependent weights in the attention layer. More importantly, however, we observe that the separation between active and passive particles becomes more pronounced at lower temperatures, which indicates some sort of increasing structural heterogeneity. This observation explains the improved performance of tGlassMLP at lower temperatures, as quantified by the Pearson correlation in Fig. 2.
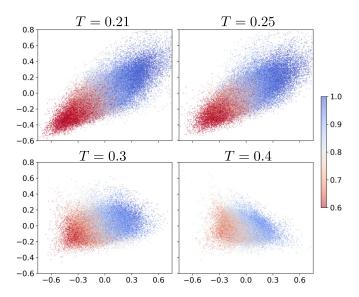

FIG. 12. Visualization of the amorphous structure, encoded by the two-dimensional bottleneck layer at fixed time, $\langle \bar{\mathcal{C}}_B \rangle = 0.8$, and different temperatures. The figure is constructed as in Fig. 11.
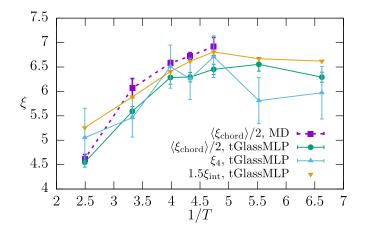
FIG. 13. Temperature-dependent length scales ξ as extracted from various techniques described in the paper at $\langle \bar{C}_B \rangle = 0.8$. The figure demonstrates a good agreement between the different lengths.

VII. DISCUSSION AND CONCLUSIONS

We presented and analyzed the tGlassMLP model, which can predict the relaxation dynamics of deeply supercooled liquids from the amorphous microstructure over a large range of time scales and temperatures. The approach has been verified by calculating Pearson correlation coefficients ρ_P with direct MD simulations and by evaluating dynamic susceptibilities χ_4 and dynamic correlation length scales ξ_4.

Overall, the predictions of tGlassMLP regarding spatial correlations support a scenario in which dynamic heterogeneities do not grow significantly when the temperature is decreased far below T_{MCT} and approaches the experimental glass transition T_g. To summarize these results, we compare in Fig. 13 the different techniques used above to calculate dynamic length scales at a time t corresponding to $\langle \bar{C}_B(t) \rangle = 0.8$. We scale the results with a method-dependent factor and thus find very good overlap between all results [55]. The figure shows that all methods produce the same trend of a dynamic length scale increasing quite rapidly at high temperatures, followed by a much slower evolution when $T < T_{MCT} = 0.3$. The length scales directly determined by MD follow the same trend, although of course on a much smaller temperature range. We also found that the evolution of the dynamic susceptibility $\chi_4(t)$ follows a trend similar to the length scales shown in Fig. 13.

Our data-driven approach to predict dynamic heterogeneity is quite general and is based on detecting structural heterogeneity, which can easily be measured using the swap Monte Carlo algorithm at very low temperatures. Nevertheless, it still leaves open the possibility that the extrapolated tGlassMLP networks do not capture all important mechanisms leading to growing heterogeneities at low temperatures. However, taking the extrapolated data at face value, it is interesting to compare the emerging scenario in which dynamic heterogeneities only grow very weakly below $T < T_{MCT}$ to available numerical and experimental data.

Several articles have featured strongly increasing dynamic heterogeneities and correlation lengths, including colloidal experiments [56] and computer simulations [47,48,57,58] in the high temperature regime. This is in agreement with our

results. The change of this behavior toward a weaker increase at temperatures lower than T_{MCT} is also found in several simulation and experimental studies. In the three-dimensional Kob-Andersen mixture a crossover from strongly increasing dynamic heterogeneities to weakly-increasing has recently been observed [59–64]. Also in other models, such as binary mixtures of quasihard spheres [65] it has been observed that dynamic heterogeneities do not continue to grow significantly. A plateau and a crossover to a weakly increasing heterogeneity has been observed in Ref. [66] for Kob-Andersen binary mixtures over a range of mixing ratios. Only very recently were measurements performed at much lower temperatures in a two-dimensional soft sphere system [50], but here the saturation was less pronounced. These above numerical findings in three dimensions are consistent with experiments on molecular glass formers near the experimental glass transition temperature, which show length scales in the order of five molecular diameters [67,68] and weakly increasing dynamic heterogeneity [69,70].

The analysis of the intrinsic properties of the trained tGlassMLP model shows that the effect of structure on dynamics can be reduced to just two variables. However, we could not find any direct interpretation for them, because they are an aggregate of several physical local features. Despite their good predicting power, there is no clear separation between mobile and immobile particles in this two-dimensional representation, thus casting doubts on the possibility of finding a connection between simple form of amorphous order or simple local structures and dynamics by unsupervised learning (similar "potato-shaped" dimensional reductions were found in various glass formers in Ref. [71]). It would be interesting to study whether this conclusion applies to the dynamics observed in individual trajectories [72,73] instead of the dynamic propensity.

Finally, we believe that the transferable ML approach presented in this paper can be an important step towards the study of dynamic heterogeneities in a manifold of glass-forming materials. This includes models in different spatial dimensions, continuously polydisperse models [50] as well as more fragile glass formers, in which dynamic heterogeneities might show a different fate at low temperatures [74]. Furthermore, quantifying the performance of transferability is an important topic in the machine learning community [75,76]. Future work could include advanced transferability measures [77] or self-supervised learning [78]. Employing such techniques to our model would be an important step to strengthen further the reliability of transferred results.

APPENDIX A: CONVERSION BETWEEN TIMES AND BOND-BREAKING CORRELATION VALUES

In Sec. II A we introduced the bond-breaking correlation and explained how the average value $\langle \bar{C}_B(t) \rangle$ is used to encode
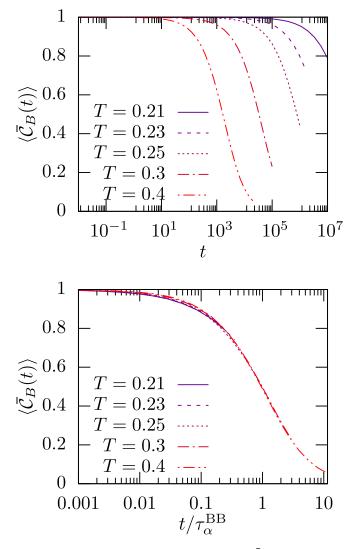
FIG. 14. Average bond-breaking correlation $\langle \bar{\mathcal{C}}_B(t) \rangle$ for different temperatures. Top and bottom show the same results, where bottom is rescaled by the estimated structural relaxation time $\tau_\alpha^{\mathrm{BB}}$.

time in the machine learning procedure. Here, we provide further evidence to support this approach.

In Fig. 14 it can be observed that the decay of $\langle \bar{\mathcal{C}}_B(t) \rangle$ at various temperatures is very similar, although the curves at different temperatures decay on different time scales. We extract the structural relaxation time $\langle \bar{\mathcal{C}}_B(t = \tau_\alpha^{\mathrm{BB}}) \rangle = 0.5$ also for the curves, which are not simulated long enough by maximizing the overlap between the curves. The result is shown in bottom panel of Fig. 14. The curves display nice time-temperature superposition, indicating that the shape of the time decay changes very little (if at all) over a wide temperature range.

Based on the extracted values for $\tau_\alpha^{\mathrm{BB}}$, which show an Arrhenius dependence above $T \leqslant 0.25$, we extrapolate the corresponding values for $\tau_\alpha^{\mathrm{BB}}$ at lower temperatures and find that the glass transition temperature $\tau_\alpha^{\mathrm{BB}}(T = T_g) = 10^{12}$ can be estimated as $T_g = 0.15$, which is the lowest temperature considered in the paper. To reconstruct the time dependence of propensity at such low temperatures, we predict propensity

at a certain value of $\langle \bar{\mathcal{C}}_B \rangle$, use Fig. 14 to estimate $t/\tau_\alpha^{\mathrm{BB}}$ and thus deduce t using the previously estimated value for $\tau_\alpha^{\mathrm{BB}}$.

APPENDIX B: DETAILS ON TGLASSMLP AND THE TRAINING PROCEDURE

The core of tGlassMLP is the M_S-dimensional input layer, followed by the attention layer, as described in Sec. II and sketched in Fig. 1. The state point is inserted into the attention layer via MLP1. MLP1 has a six-dimensional input layer, in which we feed the slightly transformed state point, $[1/T, 1/T^2, 1/T^3, \langle \bar{\mathcal{C}}_B(t) \rangle, \langle \bar{\mathcal{C}}_B(t) \rangle^2, \langle \bar{\mathcal{C}}_B(t) \rangle^3]$, and two six-dimensional hidden layers, each with an ELU activation function [79]. The two-dimensional output is then calculated using a linear activation function.

After the attention layer, the data is encoded into the two-dimensional bottleneck layer, which is concatenated with the output by MLP2, and followed by two ten-dimensional hidden layers, each with an ELU activation function. MLP2 is similar to MLP1 just formed of two seven-dimensional hidden layers and a four-dimensional output layer. The performance of tGlassMLP, however, does not crucially depend on any of these choices. The final output $\mathcal{X}_{\mathrm{MLP}}^i$ is then calculated using a linear activation function.

TABLE I. Overview over the different state points used for the training of tGlassMLP. "x" denotes training data extracted from MD simulations, "t" are transferred results using the first training step in "equi-time" mode. "r" are results shown for the final trained tGlassMLP networks.

tGlassMLP ($T_{\mathrm{min}} = 0.21$)

$\langle \bar{\mathcal{C}}_B(t) \rangle \diagdown T$	0.15	0.18	0.21	0.23	0.25	0.3	0.4
0.99	t	t	x	x	x	x	x
0.94	t	t	x	x	x	x	x
0.90	t	t	x	x	x	x	x
0.85	t	t	x	x	x	x	x
0.8	t	t	x	x	x	x	x
0.75	r	t	t	x	x	x	x
0.65	r	t	t	t	x	x	x
0.5	r	t	t	t	x	x	x
0.25	r	r	r	r	r	x	x
0.15							x
0.075							x

tGlassMLP ($T_{\mathrm{min}} = 0.25$)

$\langle \bar{\mathcal{C}}_B(t) \rangle \diagdown T$	0.21	0.23	0.25	0.3	0.4
0.99	t	t	x	x	x
0.94	t	t	x	x	x
0.90	t	t	x	x	x
0.85	t	t	x	x	x
0.8	t	t	x	x	x
0.75	t	t	x	x	x
0.65	t	t	x	x	x
0.5	t	t	x	x	x
0.25	r	r	x	x	x
0.15				x	
0.075					x

TABLE II. Accuracies of the Adam optimizer used for training. "e" stands for the number of epochs.

Phase 1	10e, 10^{-3}	50e, 5×10^{-4}	25e, 2×10^{-4}
Phase 2	50e, 4×10^{-5}		

As described in the main text, tGlassMLP is then trained in two steps: The first "equi-time" step for constant $\langle \bar{\mathcal{C}}_B(t) \rangle$ is followed by a self-consistency check and transfer procedure. The results are then used for the final training of the full tGlassMLP model. The data that is used for the tGlassMLP models presented in this paper is summarized in Table I.

Each training step is performed in a similar way as the training of GlassMLP [27]. The batch size is equal the number of type 1 particles per configuration, $N_{\text{batch}} = N_1 = 600$. The training of both the "equi-time" step and the final step is separated into two phases. In the first phase, the model is trained for 85 epochs with a loss function that only considers the mean squared error. The accuracy of the Adam optimizer is varied as described in Table II. Afterwards, the model is trained for another 50 epochs using the loss function with the same parameters as for GlassMLP.

When applying tGlassMLP on a large set of amorphous structures, in particular at low temperatures and long times, we use an iterative procedure to constrain the predicted mean propensity $\overline{\mathcal{X}_{\text{MLP}}} = \frac{1}{N_1} \sum_{i=1}^{N_1} \mathcal{X}_{\text{MLP}}^i$ to the expected value $\langle \bar{\mathcal{C}}_B(t) \rangle$. To achieve this we define $\mathcal{C}_B^0 = \langle \bar{\mathcal{C}}_B(t) \rangle$ as the initial input in tGlassMLP. Using the output in iteration j, we update the input via the recursive relation $\mathcal{C}_B^{j+1} = \mathcal{C}_B^j + 0.75(\overline{\mathcal{X}_{\text{MLP}}}^j - \langle \bar{\mathcal{C}}_B(t) \rangle)$. We iterate for a maximum of 6 iterations, and check for convergence. The final $\mathcal{C}_B^j$ is never very different from $\langle \bar{\mathcal{C}}_B(t) \rangle$ due to the self-consistency check described in the main text.
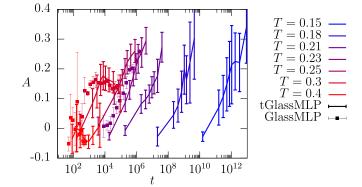

FIG. 15. Higher-order prefactor A, extracted from S_4 after fitting the length scale ξ_{S_4}.

APPENDIX C: HIGHER-ORDER TERM IN FOUR-POINT STRUCTURE FACTOR

In this Appendix we analyze the higher-order prefactor A as extracted from fitting the dynamic structure factor S_4 with an additional contribution, which is of third order in q,

$$S_4(0.2 < q < 0.6, t) = \tilde{\chi}_4(t)/(1 + (\xi q)^2 + A(\xi q)^3). \quad \text{(C1)}$$

As discussed in Ref. [27] this quantity shows a very pronounced increase around the mode-coupling temperature T_{MCT}, which we can perfectly reproduce using tGlassMLP (see Fig. 15). Now, we can also systematically study temperatures below T_{MCT}. We find that when entering deep into the glassy regime, also the prefactor A reaches a plateau, indicating that the geometry of rearranging clusters is not significantly changing anymore, consistent with Ref. [61].

[1] E.-J. Donth, *The Glass Transition: Relaxation Dynamics in Liquids and Disordered Materials* (Springer Science & Business Media, New York, 2001), Vol. 48.

[2] P. G. Debenedetti and F. H. Stillinger, Supercooled liquids and the glass transition, Nature (London) **410**, 259 (2001).

[3] L. Berthier, G. Biroli, J.-P. Bouchaud, L. Cipelletti, and W. Saarloos, *Dynamical Heterogeneities in Glasses, Colloids, and Granular Media* (Oxford University Press, Oxford, 2011), pp. 1–464.

[4] A. Widmer-Cooper, H. Perry, P. Harrowell, and D. R. Reichman, Irreversible reorganization in a supercooled liquid originates from localized soft modes, Nat. Phys. **4**, 711 (2008).

[5] C. P. Royall and S. R. Williams, The role of local structure in dynamical arrest, Phys. Rep. **560**, 1 (2015).

[6] H. Tanaka, H. Tong, R. Shi, and J. Russo, Revealing key structural features hidden in liquids and glasses, Nat. Rev. Phys. **1**, 333 (2019).

[7] D. Richard, M. Ozawa, S. Patinet, E. Stanifer, B. Shang, S. A. Ridout, B. Xu, G. Zhang, P. K. Morse, J.-L. Barrat, L. Berthier, M. L. Falk, P. Guan, A. J. Liu, K. Martens, S. Sastry, D.

Vandembroucq, E. Lerner, and M. L. Manning, Predicting plasticity in disordered solids from structural indicators, Phys. Rev. Mater. **4**, 113609 (2020).

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

[9] B. Doliwa and A. Heuer, What does the potential energy landscape tell us about the dynamics of supercooled liquids and glasses? Phys. Rev. Lett. **91**, 235501 (2003).

[10] 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).

[11] J. M. Rieser, C. P. Goodrich, A. J. Liu, and D. J. Durian, Divergence of Voronoi cell anisotropy vector: A threshold-free characterization of local structure in amorphous materials, Phys. Rev. Lett. **116**, 088001 (2016).

[12] A. Malins, J. Eggers, C. Royall, S. Williams, and H. Tanaka, Identification of long-lived clusters and their link to slow dynamics in a model glass former, J. Chem. Phys. **138**, 12A535 (2013).

[13] 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).

[14] G. M. Hocky, D. Coslovich, A. Ikeda, and D. R. Reichman, Correlation of local order with particle mobility in supercooled liquids is highly system dependent, Phys. Rev. Lett. **113**, 157801 (2014).

[15] M. Lerbinger, A. Barbot, D. Vandembroucq, and S. Patinet, Relevance of shear transformations in the relaxation of supercooled liquids, Phys. Rev. Lett. **129**, 195501 (2022).

[16] L. Berthier, Self-induced heterogeneity in deeply supercooled liquids, Phys. Rev. Lett. **127**, 088002 (2021).

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

[18] N. Oyama, S. Koyama, and T. Kawasaki, What do deep neural networks find in disordered structures of glasses? Front. Phys. **10**, 1007861 (2023).

[19] S. Soltani, C. W. Sinclair, and J. Rottler, Exploring glassy dynamics with Markov state models from graph dynamical neural networks, Phys. Rev. E **106**, 025308 (2022).

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

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

[22] R. M. Alkemade, F. Smallenburg, and L. Filion, Improving the prediction of glassy dynamics by pinpointing the local cage, J. Chem. Phys. **158**, 134512 (2023).

[23] 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).

[24] 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 (2016).

[25] Z.-Y. Yang, D. Wei, A. Zaccone, and Y.-J. Wang, Machine-learning integrated glassy defect from an intricate configurational-thermodynamic-dynamic space, Phys. Rev. B **104**, 064108 (2021).

[26] T. M. Obadiya and D. M. Sussman, Using fluid structures to encode predictions of glassy dynamics, Phys. Rev. Res. **5**, 043112 (2023).

[27] G. Jung, G. Biroli, and L. Berthier, Predicting dynamic heterogeneity in glass-forming liquids by physics-inspired machine learning, Phys. Rev. Lett. **130**, 238202 (2023).

[28] 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 (2020).

[29] F. S. Pezzicoli, G. Charpiat, and F. P. Landes, Se (3)-equivariant graph neural networks for learning glassy liquids representations, arXiv:2211.03226.

[30] H. Shiba, M. Hanai, T. Suzumura, and T. Shimokawabe, Botan: Bond targeting network for prediction of slow glassy dynamics by machine learning relative motion, J. Chem. Phys. **158**, 084503 (2023).

[31] X. Jiang, Z. Tian, K. Li, and W. Hu, A geometry-enhanced graph neural network for learning the smoothness of glassy dynamics from static structure, J. Chem. Phys. **159**, 144504 (2023).

[32] 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).

[33] S. Ciarella, M. Chiappini, E. Boattini, M. Dijkstra, and L. M. Janssen, Dynamics of supercooled liquids from static averaged quantities using machine learning, Mach. Learn. Sci. Technol. **4**, 025010 (2023).

[34] G. Zhang, H. Xiao, E. Yang, R. J. S. Ivancic, S. A. Ridout, R. A. Riggleman, D. J. Durian, and A. J. Liu, Structuro-elasto-plasticity model for large deformation of disordered solids, Phys. Rev. Res. **4**, 043026 (2022).

[35] A. Ninarello, L. Berthier, and D. Coslovich, Models and algorithms for the next generation of glass transition studies, Phys. Rev. X **7**, 021039 (2017).

[36] L. Berthier, E. Flenner, C. J. Fullerton, C. Scalliet, and M. Singh, Efficient swap algorithms for molecular dynamics simulations of equilibrium supercooled liquids, J. Stat. Mech. (2019) 064004.

[37] 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).

[38] L. Berthier, D. Coslovich, A. Ninarello, and M. Ozawa, Equilibrium sampling of hard spheres up to the jamming density and beyond, Phys. Rev. Lett. **116**, 238002 (2016).

[39] A. D. S. Parmar, M. Ozawa, and L. Berthier, Ultrastable metallic glasses in silico, Phys. Rev. Lett. **125**, 085505 (2020).

[40] 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).

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

[42] C. H. Rycroft, Voro++: A three-dimensional Voronoi cell library in C++, Chaos **19**, 041111 (2009).

[43] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process., Vol. 30 (2017).

[44] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[45] G. Jung, R. M. Alkemade, V. Bapst, D. Coslovich, L. Filion, F. P. Landes, A. Liu, F. S. Pezzicoli, H. Shiba, G. Volpe *et al.*, Roadmap on machine learning glassy liquids, arXiv:2311.14752.

[46] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.109.064205 for a video on the structural relaxation at the glass transition temperature.

[47] C. Bennemann, C. Donati, J. Baschnagel, and S. C. Glotzer, Growing range of correlated motion in a polymer melt on cooling towards the glass transition, Nature (London) **399**, 246 (1999).

[48] E. Flenner, M. Zhang, and G. Szamel, Analysis of a growing dynamic length scale in a glass-forming binary hard-sphere mixture, Phys. Rev. E **83**, 051501 (2011).

[49] L. Berthier and G. Biroli, Theoretical perspective on the glass transition and amorphous materials, Rev. Mod. Phys. **83**, 587 (2011).

[50] C. Scalliet, B. Guiselin, and L. Berthier, Thirty milliseconds in the life of a supercooled liquid, Phys. Rev. X **12**, 041028 (2022).

[51] S. Karmakar, C. Dasgupta, and S. Sastry, Analysis of dynamic heterogeneity in a glass former from the spatial correlations of mobility, Phys. Rev. Lett. **105**, 015701 (2010).

[52] S. Karmakar, C. Dasgupta, and S. Sastry, Growing length scales and their relation to timescales in glass-forming liquids, Annu. Rev. Condens. Matter Phys. **5**, 255 (2014).

[53] E. Flenner and G. Szamel, Dynamic heterogeneity in two-dimensional supercooled liquids: Comparison of bond-breaking and bond-orientational correlations, J. Stat. Mech. (2016) 074008.

[54] H. Shiba, Y. Yamada, T. Kawasaki, and K. Kim, Unveiling dimensionality dependence of glassy dynamics: 2D infinite fluctuation eclipses inherent structural relaxation, Phys. Rev. Lett. **117**, 245701 (2016).

[55] The factor of 1.5 to scale ξ_{int} is determined empirically, and depends on the absolute weights used in the MLP to encode the output of the attention layer to the bottleneck layer. The weights of the attention layer can obviously only capture relative changes of the overall length scale with temperatures. Similarly, the factor of 2 for the chord length is reasonable since chords represent the linear extension of mobile regions, and ξ_4, approximately, their radius.

[56] E. R. Weeks, J. C. Crocker, A. C. Levitt, A. Schofield, and D. A. Weitz, Three-dimensional direct imaging of structural relaxation near the colloidal glass transition, Science **287**, 627 (2000).

[57] S. Whitelam, L. Berthier, and J. P. Garrahan, Dynamic criticality in glass-forming liquids, Phys. Rev. Lett. **92**, 185705 (2004).

[58] P. Charbonneau and G. Tarjus, Decorrelation of the static and dynamic length scales in hard-sphere glass formers, Phys. Rev. E **87**, 042305 (2013).

[59] G. M. Hocky, L. Berthier, W. Kob, and D. R. Reichman, Crossovers in the dynamics of supercooled liquids probed by an amorphous wall, Phys. Rev. E **89**, 052311 (2014).

[60] D. Coslovich, M. Ozawa, and W. Kob, Dynamic and thermodynamic crossover scenarios in the Kob-Andersen mixture: Insights from multi-CPU and multi-GPU simulations, Eur. Phys. J. E **41**, 62 (2018).

[61] P. Das and S. Sastry, Crossover in dynamics in the Kob-Andersen binary mixture glass-forming liquid, J. Non-Cryst. Solids X **14**, 100098 (2022).

[62] E. Flenner and G. Szamel, Dynamic heterogeneities above and below the mode-coupling temperature: Evidence of a dynamic crossover, J. Chem. Phys. **138**, 12A523 (2013).

[63] L. Berthier, G. Biroli, J.-P. Bouchaud, W. Kob, K. Miyazaki, and D. R. Reichman, Spontaneous and induced dynamic fluctuations in glass formers. I. General results and dependence on ensemble and dynamics, J. Chem. Phys. **126**, 184503 (2007).

[64] L. Berthier, G. Biroli, J.-P. Bouchaud, W. Kob, K. Miyazaki, and D. R. Reichman, Spontaneous and induced dynamic correlations in glass formers. II. Model calculations and comparison to numerical simulations, J. Chem. Phys. **126**, 184504 (2007).

[65] W. Kob, S. Roldán-Vargas, and L. Berthier, Non-monotonic temperature evolution of dynamic correlations in glass-forming liquids, Nat. Phys. **8**, 164 (2012).

[66] L. Ortlieb, T. S. Ingebrigtsen, J. E. Hallett, F. Turci, and C. P. Royall, Probing excitations and cooperatively rearranging regions in deeply supercooled liquids, Nat. Commun. **14**, 2621 (2023).

[67] M. D. Ediger, Spatially heterogeneous dynamics in supercooled liquids, Annu. Rev. Phys. Chem. **51**, 99 (2000).

[68] L. Berthier, G. Biroli, J.-P. Bouchaud, L. Cipelletti, D. E. Masri, D. L'Hôte, F. Ladieu, and M. Pierno, Direct experimental evidence of a growing length scale accompanying the glass transition, Science **310**, 1797 (2005).

[69] C. Dalle-Ferrier, C. Thibierge, C. Alba-Simionesco, L. Berthier, G. Biroli, J.-P. Bouchaud, F. Ladieu, and D. L'Hôte, and G. Tarjus, Spatial correlations in the dynamics of glassforming liquids: Experimental determination of their temperature dependence, Phys. Rev. E **76**, 041510 (2007).

[70] S. Albert, T. Bauer, M. Michl, G. Biroli, J.-P. Bouchaud, A. Loidl, P. Lunkenheimer, R. Tourbot, C. Wiertel-Gasquet, and F. Ladieu, Fifth-order susceptibility unveils growth of thermodynamic amorphous order in glass-formers, Science **352**, 1308 (2016).

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

[72] L. O. Hedges, R. L. Jack, J. P. Garrahan, and D. Chandler, Dynamic order-disorder in atomistic models of structural glass formers, Science **323**, 1309 (2009).

[73] R. Pinchaipat, M. Campo, F. Turci, J. E. Hallett, T. Speck, and C. P. Royall, Experimental evidence for a structural-dynamical transition in trajectory space, Phys. Rev. Lett. **119**, 028004 (2017).

[74] C. P. Royall, A. Malins, A. J. Dunleavy, and R. Pinney, Strong geometric frustration in model glassformers, J. Non-Cryst. Solids **407**, 34 (2015).

[75] Y. Bao, Y. Li, S.-L. Huang, L. Zhang, L. Zheng, A. Zamir, and L. Guibas, An information-theoretic approach to transferability in task transfer learning, in *Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP)* (IEEE, Piscataway, NJ, 2019), pp. 2309–2313.

[76] A. T. Tran, C. V. Nguyen, and T. Hassner, Transferability and hardness of supervised classification tasks, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (IEEE, Piscataway, NJ, 2019), pp. 1395–1405.

[77] C. Nguyen, T. Hassner, M. Seeger, and C. Archambeau, Leep: A new measure to evaluate transferability of learned representations, in *International Conference on Machine Learning* (PMLR, Vienna, Austria, 2020), pp. 7294–7305.

[78] S. Gidaris, P. Singh, and N. Komodakis, Unsupervised representation learning by predicting image rotations, arXiv:1803.07728.

[79] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, Fast and accurate deep network learning by exponential linear units (ELUs), arXiv:1511.07289.