



HAL
open science

Roadmap on machine learning glassy liquids

Gerhard Jung, Rinske M Alkemade, Victor Bapst, Daniele Coslovich, Laura Filion, François P. Landes, Andrea Liu, Francesco Saverio Pezzicoli, Hayato Shiba, Giovanni Volpe, et al.

► **To cite this version:**

Gerhard Jung, Rinske M Alkemade, Victor Bapst, Daniele Coslovich, Laura Filion, et al.. Roadmap on machine learning glassy liquids. *Nature Rev.Phys.*, 2025, 7 (2), pp.91-104. 10.1038/s42254-024-00791-4 . hal-04344224

HAL Id: hal-04344224

<https://hal.science/hal-04344224v1>

Submitted on 5 Feb 2025

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Roadmap on machine learning glassy dynamics

Gerhard Jung,^{1,2} Rinske M. Alkemade,³ Victor Bapst,⁴ Daniele Coslovich,⁵ Laura Filion,³ François P. Landes,⁶ Andrea Liu,^{7,8} Francesco Saverio Pezzicoli,⁶ Hayato Shiba,⁹ Giovanni Volpe,¹⁰ Francesco Zamponi,¹¹ Ludovic Berthier,^{1,12} and Giulio Biroli¹¹

¹*Laboratoire Charles Coulomb (L2C), Université de Montpellier, CNRS, 34095 Montpellier, France*

²*Laboratoire Interdisciplinaire de Physique (LIPhy),
Université Grenoble Alpes, 38402 Saint-Martin-d'Hères, France*

³*Soft Condensed Matter and Biophysics, Debye Institute for Nanomaterials Science, Utrecht University, Utrecht, Netherlands*

⁴*GoogleDeepMind, London, UK*

⁵*Dipartimento di Fisica, Università di Trieste, Strada Costiera 11, 34151, Trieste, Italy*

⁶*Université Paris-Saclay, CNRS, INRIA, Laboratoire Interdisciplinaire
des Sciences du Numérique, TAU team, 91190 Gif-sur-Yvette, France*

⁷*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*

⁸*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

⁹*Graduate School of Information Science, University of Hyogo, Kobe 650-0047, Japan*

¹⁰*Department of Physics, University of Gothenburg, Origovägen 6B, Gothenburg 41296, Sweden*

¹¹*Laboratoire de Physique de l'École Normale Supérieure, ENS, Université PSL,
CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France*

¹²*Gulliver, UMR CNRS 7083, ESPCI Paris, PSL Research University, 75005 Paris, France*

(Dated: September 27, 2024)

Unraveling the connections between microscopic structure, emergent physical properties, and slow dynamics has long been a challenge when studying the glass transition. The absence of clear visible structural order in amorphous configurations complicates the identification of the key physical mechanisms underpinning slow dynamics. The difficulty in sampling equilibrated configurations at low temperatures hampers thorough numerical and theoretical investigations. This perspective article explores the potential of machine learning (ML) techniques to face these challenges, building on the algorithms that have revolutionized computer vision and image recognition. We present recent successful ML applications, as well as many open problems for the future, such as transferability and interpretability of ML approaches. We highlight new ideas and directions in which ML could provide breakthroughs to better understand the fundamental mechanisms at play in glass-forming liquids. To foster a collaborative community effort, this article also introduces the “GlassBench” dataset, providing simulation data and benchmarks for both two-dimensional and three-dimensional glass-formers. We propose critical metrics to compare the performance of emerging ML methodologies, in line with benchmarking practices in image and text recognition. The goal of this roadmap is to provide guidelines for the development of ML techniques in systems displaying slow dynamics, while inspiring new directions to improve our theoretical understanding of glassy liquids.

I. INTRODUCTION

When supercooled liquids undergo a glass transition, a dramatic slowdown of transport properties is observed and the resulting material dynamically resembles a crystalline solid — yet one of the main characteristic of glasses is that they maintain their amorphous liquid structure [1]. While glasses and other amorphous materials have been used since prehistoric times, their technological applications have blossomed in recent years [2], including, e.g., metallic glasses for biomedical implants [3] and vapor-deposited organic films used extensively in visual displays [4]. Despite several decades of research involving experiments, theory and computer simulations, many fundamental mechanisms remain to be elucidated, such as macroscopic mechanical properties, highly cooperative stress relaxation in glasses, and the statistical mechanics nature of the glass transition itself [5].

The raise of deep learning in the last decade [6] was initially driven by applications in computer vision, in particular image recognition and feature detection, which soon

outperformed traditional techniques [7]. These original breakthroughs are now starting to revolutionize several other areas in technology and science. Our aim here is to address the potential of ML methods to boost research on fundamental aspects of glassy dynamics, in particular the ones that are playing an important role in advancing theories of the glass transition.

One of the main challenges in developing a fundamental microscopic theory of glasses is the absence of any simple and visible structural order. While crystalline defects in the otherwise well-ordered structures are easily detectable, it is an open problem to find analogous structural features in amorphous materials. Over the years, many different proposals for “defects” or locally preferred structures have been proposed and developed [8–11]. This seems to indicate that, even in amorphous configurations, it could be possible to detect the emergence of some kind of short- and medium-range order. However, these identifications usually only apply to specific systems, and they are often only weakly correlated with local dynamical relaxations. There is therefore a clear need for

new and more powerful system-independent ways to systematically find preferred structures in amorphous configurations. This is a challenge for which new ML methods could be a great asset, in particular thanks to the progress in unsupervised learning. Several approaches have been developed recently towards this goal [12–15].

Another long-standing challenge has been understanding and characterizing the fundamental mechanisms underpinning slow and glassy dynamics, which are responsible for the glass transition. To this aim, there has been a substantial effort to identify the microscopic properties that lead to dynamical relaxation. Given a snapshot (an equilibrium configuration), several local properties have been proposed to pinpoint the regions more likely to relax within a window of, say, some fraction of the relaxation time. Examples include the local Debye-Waller factor, eigenvectors of the Hessian of inherent structures, etc. [16, 17]. There is no consensus on what are the best predictors of future dynamics. Moreover, they could change with temperature or be system-specific, according to several theories of the glass transition [18]. Thanks to the advances in numerical simulations of glass-forming liquids we can now produce large datasets of initial configurations and subsequent dynamical trajectories. This provides a natural playground to apply supervised learning techniques in order to identify the local predictors of dynamical relaxation. Several researchers have taken up this challenge and developed ML methods to predict where local relaxations are more likely to take place given an initial snapshot [19–30].

Finally, the ultimate goal of the research efforts devoted to the theory of glassy dynamics would be to combine the solutions of the previous problems to develop an effective theory of the glass transition. Until now, this challenge has been tackled starting from some theoretical assumptions driven by experimental and simulation results [18]. Also in this challenge ML methods can make a difference; they can assist in this quest by providing a complementary identification of the mechanisms inducing relaxation, as first shown in [31].

Clearly, the time is ripe for investigating the ability of the new ML methods to advance our fundamental understanding of glass-forming liquids. In this roadmap, focusing on the three main goals described above, we present respectively in Secs. II, III, IV the recent contributions in this endeavor, discussing the main difficulties ahead and the possible paths to circumvent them. In Fig. 1, we give a visual overview over the different ML concepts that will be discussed within the individual sections. In Sec. V we provide a framework “GlassBench” to enable, encourage and structure a broader community effort to further develop such ML approaches. It consists in (i) a dataset including simulation data for a two-dimensional [30] and a three-dimensional glass-former [25, 32], (ii) benchmarks on different tasks associated to predicting local dynamics from a given initial configuration, (iii) an assessment of the current state of the art. Our purpose is to fuel and organize new developments of advanced ML techniques,



FIG. 1. Visual summary of the roadmap. The individual sections at the center are connected to the big questions in the field of glass physics. They are surrounded by the various machine learning concepts used to answer them.

as done in the field of image and text recognition, as well as generative modelling [33]. Finally, we close this roadmap by discussing in Sec. VI exciting new concepts and directions in ML which have the potential to play a very important role in future research on the theory of glassy dynamics.

II. MACHINE LEARNING LOCALLY FAVORED STRUCTURES

Although glass-forming liquids and glasses lack long-range order, close inspection of their atomic structure reveals particle arrangements that are more regular, symmetric, and of lower (free-)energy than the average. Icosahedral local structures are the best known example of such favorable arrangements and they are found in several metallic alloys, colloidal suspensions as well as in computer models of glassy liquids [8]. Such locally favored structures (LFS), distinct from the bulk of the particle arrangements and yet incompatible with crystalline order, are also key ingredients of some theoretical approaches to glass formation [18, 34].

Despite the importance of structural analysis in glassy materials [8, 9], there is at present no generally accepted operational definition of LFS. Standard approaches, such as Voronoi tessellation [35], topological cluster classifi-

cation [36] and other related methods [37, 38], provide a detailed classification of the possible local geometric arrangements. These methods may indicate what are the most frequent or most stable local arrangements, but they are sensitive to thermal fluctuations and tend to provide a too fine-grained classification, which is difficult to exploit in a theoretical setting. Bond-order parameters (BOP) [39] provide yet another route to characterize the local structure of dense particle systems [9]. While this approach offers in principle a systematic description of the local arrangements, the choice of the relevant BOP has been traditionally guided by physical intuition [9], which requires specific and system-dependent a priori knowledge about the relevant symmetries of the local arrangements.

Unsupervised learning methods offer natural system-independent ways to tackle the above issues [40–42]. Along with automated identification of phase transitions [43–45], characterization of the properties of complex materials from high-dimensional datasets is indeed one of the key applications of unsupervised learning in condensed matter physics [41]. The general idea is to first characterize a faithful, high-dimensional representation of the particle local environment based for instance on a systematic bond-order expansion of the local density [46] (see Ref. [47] for a recent review on structural descriptors). Unsupervised ML methods are then used to identify a small number of collective coordinates, \tilde{X}_i , that account for the relevant fluctuations of the local structure, thereby reducing the dimensionality of the descriptors. Dimensionality reduction techniques range from simple principal component analysis (PCA), or its kernel variant, to more sophisticated statistical learning methods, like neural network auto-encoders (AE) [40]. These methods may be combined in the future with more advanced approaches, such as self-supervised learning or pre-training (see Sec. VIB), possibly exploiting the intrinsic symmetries of the system [48]. Clustering methods can be finally applied to this reduced structural representation of the material structure, to pinpoint its heterogeneity [41].

The studies highlighted in Ref. [41] focus mostly on ordered materials or disordered systems with covalent or hydrogen bonding, such as amorphous carbon [49] or liquid water [50], where the preferred geometrical order is easy to identify thanks to low coordination numbers. Dense amorphous systems are characterized instead by close-packed arrangements, which provide a challenging benchmark for this kind of structural analysis. In a series of recent papers [12, 13, 51–53], dimensionality reduction and clustering have been applied to models of closed-packed glass-forming liquids. In particular, Boattini *et al.* [12] have used BOP, Gaussian mixture models and a neural network AE to study glassy binary mixtures. Their work revealed a significant structural heterogeneity, suggesting that in these systems, we can distinguish fluctuating regions displaying two different types of local disorder. These regions display a connection with the

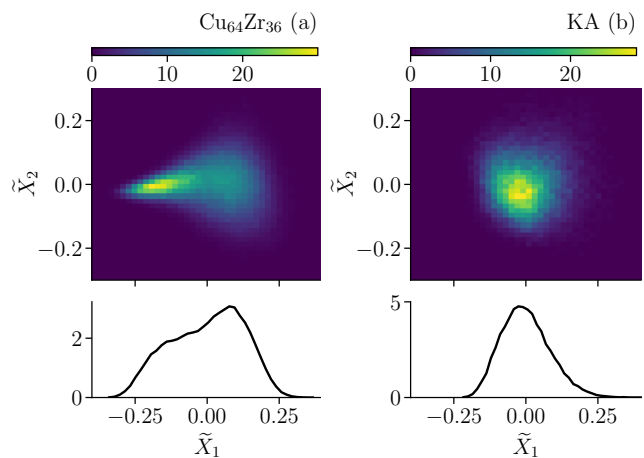


FIG. 2. Principal component analysis maps of the first two principal component projections \tilde{X}_1 and \tilde{X}_2 of the smooth bond-order parameter around (a) Cu atoms in glassy $\text{Cu}_{64}\text{Zr}_{36}$ and (b) around small particles in the Kob-Andersen binary mixture around its mode-coupling temperature [51]. Note the contrast between the bimodal structure in (a) due to icosahedral local structures [51] and the homogeneous distribution of the projections in (b). The bottom panels show the marginal distribution of the first principal component projection.

temporal fluctuations of dynamics, which will be further discussed in Sec. III.

A related study by Paret *et al.* [13] addressed the issue of clustering of local structural arrangements using a different, information-theoretic approach. At a qualitative level, the results of Refs. [12] and [13] appear consistent with one another. However, a more recent investigation [51] revealed a significant system-dependence of structural heterogeneity in glassy liquids. The gist of these findings is illustrated in Fig. 2, which shows representative PCA maps obtained from a smooth bond-order (SBO) descriptor (see supplemental material (SM) for technical details [54]). The distribution of the first two principal components is bimodal for an embedded-atom model of $\text{Cu}_{64}\text{Zr}_{36}$, which has a well-defined icosahedral LFS, while it is less heterogeneous for the canonical Kob-Andersen (KA) mixture (see also Sec. V), whose local arrangements display clearly different geometrical states. While these differences question the universality of the concept of LFS, the first few principal component projections always correlate with physically-motivated structural measures [51]. We expect that more information could be harvested by looking at chemically-resolved descriptors [55] and on larger length scales (medium range order). Moreover, computing the intrinsic dimension of structural datasets [45, 56] may provide additional insight into the nature of structural order and its system dependence.

One of the striking observations of Ref. [51] is that neural network AE and PCA yield identical reductions of the BOP descriptors. This suggests that either the local

structure in dense glassy mixtures is simple: it displays a broad continuous spectrum of geometrical arrangements, possibly decorated by features like LFS, crystallites or structural defects, or that the current identification of LFS is missing some crucial ingredient (e.g. information about dynamics). The outcome of PCA is also easy to interpret, as the principal component directions provide direct insight into the dominant structural parameters [57]. On the one hand, these results question the utility of complex deep learning methods in studying glass structure. On the other hand, the descriptors used in Refs. [12, 13, 51] do not exhaust all forms of structural heterogeneity and some may be affected by some deeper shortcomings [58]. Development of structural descriptors is still very active [59, 60] and these advances wait for applications in the context of glassy materials. Addressing the above issues may become crucial in future studies of more demanding benchmarks for structural characterization, such as compositional order in polydisperse glassy models [61, 62], medium-range order in oxides or metallic glasses [63, 64], and orientational order in glassy water [65, 66]. Computational studies of these complex systems represent exciting opportunities to gain insight into the nature and role of local structure in glassy materials and to provide solid grounds for predictive theoretical approaches based on structure.

Another research line where structure-based ML approaches are making progress aims at predicting macroscopic properties of glasses relevant for applications, such as oxides or chalcogenide glasses, over a wide range of chemical compositions [67, 68]. Recent work on sodium-silicate glasses shows that physics-informed machine learning models can reliably interpolate and extrapolate these properties on the basis of structural information only [69]. These findings indicate that, despite the apparent complexity of the feature space, the relationship between local structure and macroscopic glass properties is often linear, which makes it easy for machine learning models to generalize outside their training set. See Ref. [70] for a recent roadmap covering this topic.

Having characterized amorphous structure, a crucial question is whether the structural descriptors are connected to emergent relaxation dynamics in the glass-forming liquid [71–74]. As will be clear in Sec. V, current unsupervised methods provide only limited insights into the heterogeneity of the dynamics, except in specific systems dominated by strong icosahedral order [12, 13, 51]. Whether this is a technical limitation of the unsupervised methods used to date or rather an intrinsic feature of supercooled dynamics remains to be clarified.

III. PREDICTION OF STRUCTURAL RELAXATION AND DYNAMIC HETEROGENEITIES

One of the central challenges for both computational and theoretical studies of glass-forming liquids is to pre-

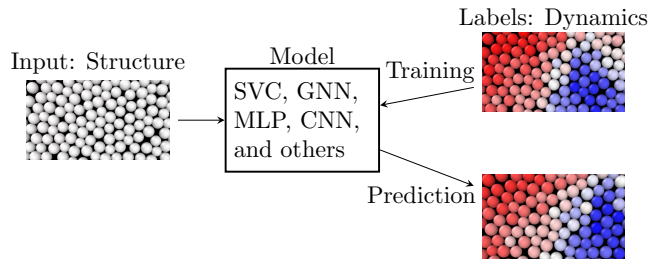


FIG. 3. Typical supervised ML procedure in condensed matter. The raw input is encoded using structural descriptors or graphs. A model is trained using labels which describe structural relaxation obtained, e.g., from molecular dynamics simulations. Colors indicate frozen (blue) or rearranging (red) particles. ML techniques range from support vector classification (SVC) to graph neural networks (GNNs), multi-layer perceptrons (MLPs), and convolutional neural networks (CNNs). After training, structural relaxation is predicted for previously unseen structures.

dict future dynamics of a configuration from an initial snapshot. Note that one is not interested in predicting the whole future evolution, but only the dynamical processes leading to microscopic irreversible motion. Supervised ML provides a natural tool to perform such prediction, essentially by fitting high-dimensional structural input to the relaxation dynamics, similar in spirit to classification in image recognition. In general, three choices need to be made to design a model: i) which structural descriptors to use to characterize the input configuration; ii) which labels to use to quantify structural microscopic relaxation; and iii) what model and ML algorithm to use to fit the input to the labels. A large variety of techniques have already been introduced to tackle this problem, ranging from ridge regression using complex and coarse-grained structural descriptors to graph neural networks using raw particles positions, as illustrated in Fig. 3.

In 2015, Cubuk *et al.* [20] demonstrated that support vector classifiers (SVC) [75] could be used to classify soft spots relaxing fast against slowly-relaxing regions in glasses. Here, soft spots are defined as regions that had a high likelihood of rearranging within a short time scale. As input to this algorithm, each particle was assigned a vector of local structural descriptors that captured the local density and angular structure within shells at different distances from its center. From the trained SVC a continuous structural descriptor called softness, S , can be extracted which correlates with the likelihood for the particle to rearrange in the near future. Softness has been used to gain insight into a variety of glass problems encompassing many different types of glassy liquids and disordered solids, ranging from strong to fragile and ductile to brittle with constituent particles ranging from atomic to granular, studied in bulk and in thin films [19, 76–84]. Further, this approach also led to a series of papers [31, 85, 86] aiming at the construction of an

effective model for the dynamics in glassy fluids built around the evolution of the softness field, which will be discussed in more detail in Sec. IV.

Whereas softness is associated to prediction on short time scales, several subsequent works have focused on predicting the dynamics, or more precisely the dynamic propensity [87], on longer time scales $t \sim \tau_\alpha$. Here, τ_α is the typical structural relaxation time scale on which each particle moves on average roughly one particle diameter (see SM [54]). Dynamic propensity quantifies the local dynamics of a glass-forming liquid by capturing the typical behaviour of each particle in a structure using the isoconfigurational ensemble [87]. This ensemble is formed by a set of trajectories which start from the same initial structure but with different initial velocities drawn from a Maxwell-Boltzmann distribution. By calculating the mean distance travelled by a given particle in this ensemble, one arrives at the dynamic propensity $\langle |\Delta \mathbf{r}_i(t)| \rangle_{\text{iso}}$. Due to this averaging, the dynamic propensity captures the part of the dynamics encoded in the initial structure, leaving out the part stemming from the initial velocities which cannot be captured by any structural descriptor. While some dynamical information is necessarily lost [88], the dynamic propensity, which fluctuates from one particle to the other, is an important measure for dynamic heterogeneity in supercooled liquids. At the end of the section, we will also discuss ways to reintroduce the fluctuations around the isoconfigurational average using different labels or new ML designs.

In 2020, Bapst *et al.* [21] introduced a graph neural network (GNN) that could predict the dynamic propensity significantly better than a support vector machine (SVM) (where the input for the SVM was angular and radial functions similar to Ref. [20]). In contrast to the SVM, the input to the GNN was a graph structure where each particle in the initial configuration was represented by a vertex, and edges were drawn between particles within a cutoff radius of each other. In addition to the structure of this graph, the vertices and edges also carry information: the particle species (as vertex data) and the vectors connecting neighbours (as edge data). The GNN model then consists of several multilayer perceptrons (MLP) that iteratively update the features contained at the edges and nodes, with each iteration passing information along the nodes and edges of the network. After the final iteration, the features at the nodes are passed through a final MLP that predicts the dynamic propensity. As all vertices are updated in parallel, the network predicts mobilities of all particles simultaneously. This network was then optimized to minimize the squared difference between the predicted and true propensity using the \mathcal{L}_2 norm. Recently, two improvements to this GNN approach were proposed. Shiba *et al.* [25] demonstrated that a significantly higher accuracy could be reached over nearly all time scales by not only considering single particle dynamics but also pairwise dynamics. Specifically, they trained the GNN to predict not only the dynamic propensity but also the iso-

configurational change in the distance between pairs of particles sharing an edge in the graph — a modification called BOnd TArgeting Network (BOTAN) [25]. Intriguingly, even with the same overall architecture, BOTAN finds a better prediction for the single particle dynamics, showing that the extra edge information improves the performance of the GNN. Pezzicoli *et al.* explored an alternative improvement by explicitly requiring the GNN to enforce the rotational symmetry, an idea sometimes referred to as geometric deep learning or rotation equivariant network (SE(3), [28]). This adaptation also improved on the original work of Ref. [21] over nearly all considered time scales.

In addition to the development of increasingly sophisticated ML methods to predict the dynamic propensity, significant efforts have been made to better capture important local features of the structural input. Shortly after the seminal work on GNNs [21], Boattini *et al.* incorporated the recursive updating properties of GNNs into a set of locally coarse-grained structural descriptors. They found that fitting just three generations of descriptors with a linear regression algorithm was sufficient to essentially reach the accuracy of the GNN — leading to a much simpler and more interpretable algorithm for fitting glassy dynamics. Interestingly, learning the dynamic propensity using these descriptors with non-linear models (MLP and GNN) did not improve the ability to predict the dynamics, as quantified by the Pearson correlation coefficient [24].

This approach has been further developed by including physics-inspired descriptors which have been identified as important structural proxies in the past thirty years of glass research (GlassMLP, [30]). These additional structural descriptors include potential energy and properties of the Voronoi cells, as well as the choice of describing the system in terms of its inherent state [30]. The inherent state corresponds to the energy minimum configuration that is closest to the actual input structure. These modifications improve the performance of the network. GlassMLP further uses MLP for supervised learning, which enables the precise representation of non-linear or non-Gaussian features such as probability distributions of propensities [30]. Using transferability in system size, the network has been applied to determine dynamic correlation lengths and the geometry of rearranging clusters over a wide range of temperatures. The GlassMLP model was recently enhanced to improve the transferability across time scales and temperatures and to explore physical regimes where direct training cannot be performed [89]. In a vein similar to GlassMLP, structural descriptors have been improved by going beyond inherent states and using cage states (CAGE, [26]). Cage states are extracted from restricted ensemble averages of the local structure using Monte Carlo simulations and thus better describe the local environment, which also helps improving the performance of the model.

By construction, all methods perform best at the temperature at which they are trained. However, it is pos-

sible to apply a trained network to other temperatures, and test how predictions correlate with true dynamics there [20, 21, 28, 83, 89]. Good performance in such transfer experiments indicates that the model captures relevant universal features in the structure-dynamics relationship of glass-forming liquids. Transferability has, for example, been used in Ref. [83] to investigate links between amorphous structure and fragility and in Ref. [89] to predict features of dynamic heterogeneity for temperatures comparable to the experimental glass transition temperature. In Sec. V, we show additional transferability experiments for several models.

In future work, it will be interesting to further exploit the transferability of supervised ML techniques to robustly extract information on structural relaxation at very low temperatures. One particular goal would concern the evolution of dynamic heterogeneity upon approaching the glass transition, and make the connection with experimental results [90]. One possible strategy to improve transferability of trained models to low-temperature regimes in which dynamics cannot be run for long enough (i.e., no or little labeled data is available) would be to use self-supervised learning [91]. More generally, self-supervised learning could also be used to enhance performance of the deep approaches (see also Sec. VIB).

To better capture the physical phenomena underpinning glassy dynamics, it seems necessary that the ML models not only faithfully predict the dynamics at the single particle level, but also correctly reproduce all statistical features of the propensity field, including spatial and temporal correlations, such as those measured by the four-point susceptibility $\chi_4(t)$ [92]. For GNNs, a known problem is over-smoothing, i.e., the predicted propensity field tends to be smoother than the ground truth. Jung *et al.* [30] introduced additional correlations-related terms in the loss function to avoid over-smoothing, and could show that the predictions indeed display more realistic correlation functions, even in a rather simple MLP architecture. An open direction for future works is the development of such improved loss functions, which could also be used in deep architectures.

A related idea is to learn not only the isoconfigurational average of the displacement, $\langle \Delta \mathbf{r}_i(t) \rangle_{\text{iso}}$, but its full statistical distribution, $P_{\text{iso}}(\Delta \mathbf{r}_i(t))$. This could be accomplished either by additionally fitting higher moments of the distribution, or by taking a generative model approach, in which one would attempt to generate realistic single-instance dynamical fields (not averaged in the isoconfigurational ensemble). This task could be achieved using a variational auto-encoder approach [93]. The spatial correlations of the isoconfigurational average being a priori different from the spatial correlations of single instances, the hope is that such generative models would reproduce these statistics more faithfully than conventional ones. To train such models, one should probably use single-instance configurations, which can also be used to regress the average. Finally, this approach might then

be usable to propose new configurations on the structural relaxation time scale τ_α , with the long-term goal to be able to develop a Monte Carlo algorithm which completely avoids the critical slowing down of the dynamics on the approach to the glass transition. Similarly, the predictions could be used to create ultra-stable glasses by generating prototypical hard neighborhoods and remove structural “defects” [94]. In Sec. VIC we discuss in more detail generative models and their applications for sampling low-temperature glassy structures.

In physics, explaining complex behavior builds on the ability of theories and models to substantially compress the inherent information of a natural phenomenon [95]. From this standpoint, large machine learning models, involving several hundreds or thousands of directly fitted parameters, do not qualify as a physical theory in the traditional sense of the term. Moreover, due to their non-linearity, neural-networks models are still rather difficult to interpret, although some progress is being made [96, 97]. This does not mean, however, that large “black-box” ML models are useless in this context: their predictions can be instrumental as part of an heuristic process, eventually leading to a simple solution to an outstanding problem [98]. This is nicely demonstrated by the results presented in this section: building on the insight of Ref. [21], Filion and coworkers achieved about the same prediction accuracy as graph neural networks using a much simpler and transparent linear regression method [24]. Linear models thus retain a strong appeal for fundamental research in glass physics, because of their simplicity and their direct mapping to the underlying structural descriptor. Interpreting the outcome of machine learning models also hinges on the ability of identifying the most relevant features – a process known as “feature selection”. Analysis of the so-called “information imbalance” has emerged as a general and elegant approach to feature selection [99]. This method has been very recently applied to identify the most relevant structural features for glassy dynamics [100].

Finally, a natural aim for future investigations is to enlarge the ML studies described above to encompass diverse glass-forming materials with complex dynamics, including active glasses as model for biological tissues [101, 102]. Although it is generally accepted that equilibrium microscopic dynamics do not influence long-time structural relaxation it is unclear which signature activity plays on structural descriptors in active glasses [103]. Similarly, very little information exists on the dynamical properties of glasses during aging [104], and whether similarly strong structure-dynamics relationships can be found as for equilibrium relaxation.

IV. MACHINE LEARNING PHENOMENOLOGICAL GLASS MODELS

In this section, we demonstrate the potential of combining ML methods with physical insights to develop ef-

fective models and phenomenological theories of slow and glassy dynamics. The starting point are the ML methods, already described in Sec. III, that identify the local structure responsible for local relaxations on short time scales. The aim is now to use this structural field as a building block to construct a phenomenological model for how structural relaxation proceeds. This approach is built upon two useful classes of models to understand dynamics in glass-forming liquids and amorphous solids subjected to mechanical strain, namely trap and elasto-plastic models.

Trap models start from the high-temperature liquid, describing the system as a distribution of energy barriers for rearrangements. A recent implementation [105] adds a facilitation mechanism: in response to a rearrangement, all the energy barriers are subject to a small random drift. This facilitated trap model has been used to explain the emergence of excess wings in the relaxation spectra [106]. Elasto-plastic models start instead from the low-temperature solid, describing the system in terms of thermally-induced rearrangement events that can trigger other rearrangements via long-ranged strain fields [107–109]. Both of these approaches have given insights into glassy dynamics, and include facilitation effects in some manner, but both describe the local structure of the liquid in a very coarse-grained, simplistic manner. As a result, both classes of models must make ad hoc assumptions on the nature of the local relaxation events. In the case of the trap model, facilitation is assumed to lead to a shift of energy barriers, while in elasto-plastic models the distribution of yield stress is imposed.

A recent approach unites the trap and elasto-plastic models by extending them to include the local structure in the form of a machine-learned microscopic structural descriptor, the softness S , as introduced in Sec. III. Softness is sufficiently accurate that the probability of rearranging for particles of a given softness S , $P_R(S)$, has an Arrhenius temperature dependence: $P(R|S) = \exp[-\Delta E(S)/T + \Delta \Sigma(S)]$, where ΔE is the energy barrier and $\Delta \Sigma$ is the entropic one, $\Delta F = \Delta E - T\Delta \Sigma$. This suggests that particles of softness S have a well-defined free energy barrier for rearrangements, $\Delta F(S)$. The spatial variation of the softness then leads to a free-energy barrier field that couples to the stress and strain fields. Note that any local structural measure that predicts rearrangements or local yield stress [110] (whether extracted using ML, as described here or in other sections of this paper, or by other means [11]) could in principle be exploited in a similar manner to construct phenomenological models of glassy dynamics or plasticity.

Thanks to this ability to extract a free energy barrier estimate $\Delta F(S)$, the softness can naturally be used to construct a phenomenological trap model. However, softness allows one to go further by considering spatial correlations. When a rearrangement occurs, it alters the softness of rearranging particles as well as that of nearby particles through near-field facilitation [111]. It also alters the softness of more distant particles by creating a

strain field that decays away from the rearrangement. Because the strain field changes the local structural environment of particles, it alters their softness. This far-field form of facilitation is well-captured by elasto-plastic models [108].

There is an interplay between rearrangements (strain), changes in softness and elasticity, with each one affecting the other two. A systematic approach to disentangling all of these effects has been introduced in Ref. [85] and implemented in a lattice structuro-elasto-plasticity model [31] for athermal systems under load. It has been applied successfully to a number of systems of varying ductility [31, 86] and used to extract insights into the microscopic factors that control ductility, such as the strength of near-field facilitation [86].

These results pave the way towards models of structural relaxation dynamics in glassy liquids. A simple trap model built upon the barriers $\Delta F(S)$ and assuming an underlying distribution of softness, $\rho(S)$, as in Ref. [112], was developed in Ref. [113]. One can also construct a version of the facilitated trap model of Ref. [105] that incorporates S . Generalising such models to supercooled liquids, however, is more challenging since one must include time-reversal invariance. Above the mode-coupling temperature, T_c , near-field facilitation should be sufficient. The hypotheses that $\rho(S)$ is nearly Gaussian [19] and that the near-field distribution of the change in softness $\Delta S(r)$ due to a rearrangement at the origin is also nearly Gaussian, are important simplifications that allows formulation of a closed theory [114]. For systems below T_c , however, it has been shown that long-ranged facilitation via strain occurs [111]. The inclusion of time-reversal invariance in a model such as a thermal elasto-plastic model [108] or structuro-elasto-plasticity model with long-ranged facilitation is a challenging open problem that needs to be solved. A precise predictor of future dynamics, as discussed in Sec. III, would be a very useful tool to adjust such models. Additionally, it would be interesting to explore whether the free energy barriers ΔF for local relaxation, extracted from the amorphous structure using softness, could be learned more directly and used to improve effective glass models.

Another path forward is to switch from a field picture to a defect picture. Most predictors of rearrangements yield particle-based quantities that are easily converted to fields, but highlight localised regions that are susceptible to rearrangement [11]. These regions can be viewed as structural defects that interact with each other and are created and destroyed by strain and rearrangements. ML could be used to learn these interactions and rules, to help build defect theories of plasticity and glassy dynamics.

In this respect, a direction that has been recently explored consists in letting a glass-forming liquid evolve via the usual thermal motion (e.g., using molecular dynamics). During such exploration, the energy is periodically minimized in order to extract a library of mechanically stable zero-temperature configurations (inherent struc-

tures). The idea is then to use ML to classify pairs of inherent structures, instead of single configurations, and check whether the pair is connected by a low-energy excitation corresponding to a localised structural defect. For example, in Ref. [115], a large library of inherent structures was constructed by such an exploration at very low temperature inside a glass basin. By means of supervised learning techniques, it was possible to train a machine that inputs a pair of inherent structures and outputs, with good precision, the classical energy barrier separating them. This allowed for a speed up in the search for glass defects by more than one order of magnitude, which is significant given the complexity of these kind of calculations.

While this preliminary study was mostly focused on very low-energy defects that are associated with thermodynamic and transport anomalies of the glass at cryogenic temperatures [115], it should in principle be straightforward to extend the techniques to detect other kinds of defects, such as those associated with plastic events under shear [11, 116] of relaxation events under equilibrium thermal motion [117]. This is a promising direction for further research.

V. PERFORMANCE METRICS AND BENCHMARKING

An essential ingredient to fuel further development of ML techniques are detailed benchmarks for existing datasets, which may allow every researcher to independently develop and test new ML approaches without complex production and preprocessing of data [33]. We provide such benchmarks for ML glass-forming liquids for different systems, different dynamical observables, and different metrics.

The benchmarks are based on the dataset ‘‘GlassBench’’ that we have created and made publicly available (<https://doi.org/10.5281/zenodo.10118191>). The whole dataset is separated into a training set which, as the name suggests, can be used to train the neural network, and a test set which should be used only for benchmarking. In addition to initial amorphous structures and trajectories, we provide precalculated dynamical descriptors and propensities, as introduced in Sec. III. We have also uploaded sample python code for reading and processing. Additional technical information on the data format is provided with the dataset.

The tasks identified for ‘‘GlassBench’’ are directly related to the open questions highlighted in the introduction: (A) Train a model to predict single particle propensity purely from structural properties. Accuracy will be quantified using the Pearson correlation coefficient. Higher accuracy in the prediction indicates that the learned structural descriptor is indeed an important precursor for future relaxation but it can also become essential when using the model to generate new configurations. (B) Train a transferable model such that it can

be accurately applied to different temperatures. This is an important task to enable investigation of structural relaxation at temperatures that are unreachable for numerical simulations. (C) Train a model which correctly predicts spatial dynamic heterogeneity (DH) as quantified by the dynamic susceptibility. The length scale of DH is growing with decreasing temperature such that at very low temperatures some regions in the system actively rearrange while others are completely frozen. DH are not only important for properties of glass-forming materials but are also core to fundamental theories of the glass transition [90, 118].

The ML techniques used for the benchmarking are summarized in Tab. I. In the following, we will refer to them just as models. A large variety of different models is represented, with very different numbers of fitting parameters and training time. Furthermore, the models use various ways to physically preprocess the structural input, either by using inherent states [30], or even by performing a Monte Carlo averaging of local cages [26]. These different factors, combined with the benchmarking provided below, should help choosing the most suitable method for a given purpose, with focus on either the highest-scoring predictions, computational efficiency, or interpretability. In addition to these ML techniques, we also include the performance of traditional structural descriptors based on physical intuition [9, 30, 119]. The model of glass-forming liquid that we selected to develop GlassBench is the very popular Kob-Andersen (KA) mixture, that we study both in two and in three dimensions.

Task A. The aim for the models is to learn correlations between the amorphous structure and the dynamic propensity of displacements, $\mathcal{R}_i(t) \equiv \langle \Delta \mathbf{r}_i(t) \rangle_{\text{iso}}$, as introduced in Sec. III. A common metric used to assess the performance of different techniques is the Pearson correlation coefficient,

$$\rho_P = \frac{\text{cov}(\mathcal{R}_i^{\text{MD}}, \mathcal{X}_i^{\text{ML}})}{\sqrt{\text{var}(\mathcal{R}_i^{\text{MD}})\text{var}(\mathcal{X}_i^{\text{ML}})}}, \quad (1)$$

between the labels $\mathcal{R}_i^{\text{MD}}$ for each particle i of type 1 in the entire dataset as obtained from molecular dynamics (MD) simulations, and the ML output $\mathcal{X}_i^{\text{ML}}$ [120]. The results are shown in Fig. 4a, with full symbols corresponding to supervised ML techniques (introduced in Sec. III) and open symbols to unsupervised techniques or physically-motivated structural descriptors (Sec. II). All technical details are provided in the SM [54]. From this figure, it can be concluded that supervised techniques nearly approach the maximal achievable correlation over the entire range of time scales. CAGE performs best for shorter times, which appears reasonable as it uses extensive Monte-Carlo simulations to characterize the local cage structure at short times. For longer times, the SE(3) GNN extension has the strongest correlation, closely trailed by the other advanced techniques. We can also observe that there is a pronounced gap between the supervised and the unsupervised techniques, indicating

	Training	ML approach	free parameters	states	training time	training hardware	app. time
BOTAN [25]	supervised	GNN	54200	th.	~ hours	NVIDIA A100	~ s
CAGE [26]	supervised	Ridge regression	2775	th.+cage	~ min	CPU	~ hours
GlassMLP [30]	supervised	MLP	615	th.+inh.	~ min	CPU	~ s
SE(3) [28]	supervised	GNN	52660	th.+inh.	~ hours	NVIDIA Tesla V100	~ s
SBO [51]	unsupervised	PCA	0	th.	–	CPU	~ s

TABLE I. Overview of the different techniques benchmarked in this roadmap. The column “states” refers to the usage of thermal (th.), inherent (inh.) or cage states. The training time corresponds to the time required to train one model at a specific temperature and time for the KA system in three dimensions. The application time corresponds to the time required to calculate the prediction for a single configuration, including preparation such as calculation of inh./cage states.

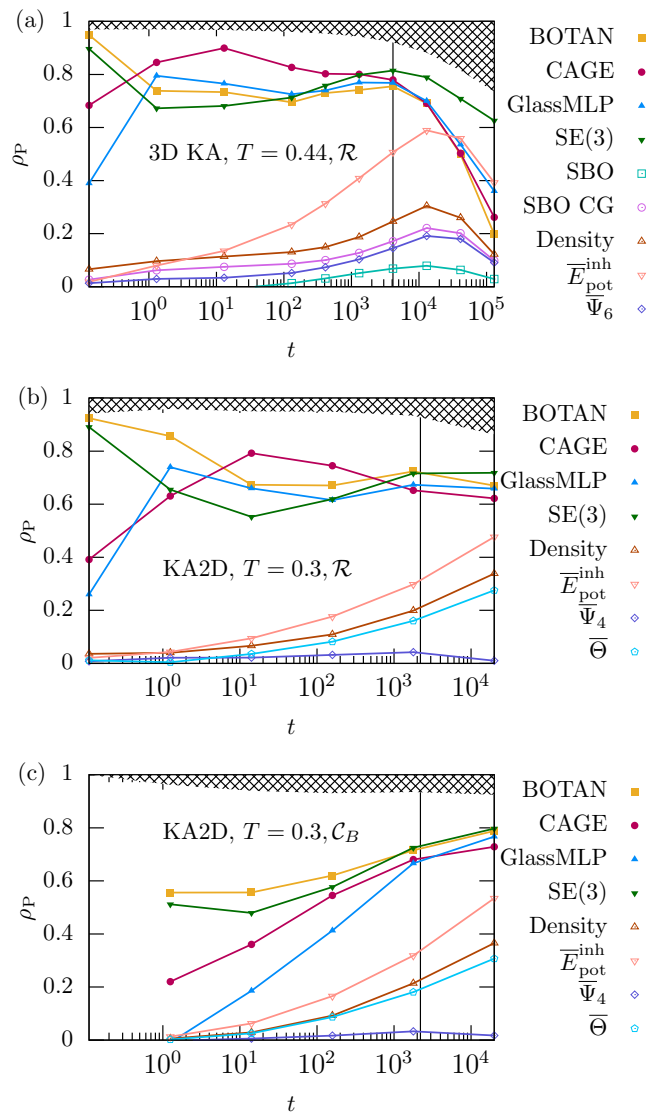


FIG. 4. **Benchmark task (A)** Pearson correlation ρ_P between various structural indicators with the ground truth. (a) 3D KA model at $T = 0.44$, (b,c) KA2D system at $T = 0.3$. The dynamical variable is (a,b) the propensity of displacements \mathcal{R} and (c) the bond-breaking propensity \mathcal{C}_B . The vertical line marks the structural relaxation time τ_α , the typical time scale on which particles rearrange, as defined in Eq. (1) in the SM [54]. The exclusion zone on the top marks the highest achievable correlation given the finite number of replicas.

that the amorphous structural features that are predictive for dynamics do not stand out in a purely structural analysis, as discussed in Sec. II.

To assess the generality of these findings, we also provide benchmarks for a two-dimensional ternary mixture of Lennard-Jones particles (KA2D) in Fig. 4b. We find that, apart from minor differences, the performance of the individual techniques is very similar to the 3D system. The most noteworthy difference is perhaps that BOTAN performs best of all methods at the structural relaxation time $t \approx \tau_\alpha$ while in the 3D KA system, the trend is reversed. While there might be subtle differences between structural relaxation across spatial dimensions the above observation implies that the problem of learning correlations between structure and dynamics is essentially independent of the spatial dimension.

The propensity of displacements \mathcal{R}_i is only one specific choice to characterise relaxation dynamics among many others. One alternative is to use the bond-breaking correlation, \mathcal{C}_B^i , which quantifies how many nearest neighbours particle i has lost during the relaxation process [30, 117]. As shown in Fig. 4c, the models also successfully learn correlations between the bond-breaking propensity and the amorphous structure. The performance of the models in the short-time predictions, however, is significantly reduced compared to the propensity of displacements. This implies that predicting the exact nature and position of the first rearrangement events appears more difficult than simply predicting short-time displacements. Around the structural relaxation time τ_α and beyond, the correlations shown in Fig. 4c for \mathcal{C}_B^i are stronger than for \mathcal{R}_i in Fig. 4b. This seemingly surprising result is connected to the growing dynamic heterogeneity at longer times (see also Fig. 6), which simplifies the prediction of larger rearranging clusters from coarse-grained structural properties, as discussed in Ref. [30]. Additionally, at times $t \geq \tau_\alpha$, we observe that the propensity of displacement \mathcal{R}_i has slowly decaying tails which are not captured by the models (see Fig. S2 in the SM) and thus likely reduce the Pearson correlation.

Task B. An important property of supervised ML techniques is their transferability, in particular towards lower temperatures, with the goal to predict the dynamics at very low temperatures that are inaccessible by direct computer simulations [28, 89]. In Fig. 5, we evaluate the capabilities of the models to transfer their structure-

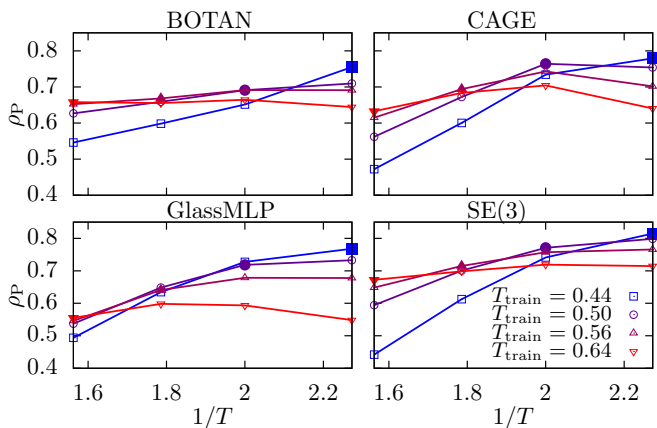


FIG. 5. **Benchmark task (B)** Transferability in temperature T of trained networks in the KA system (3D) at the structural relaxation time τ_α . Each network is trained at T_{train} (marked by the full symbol) and applied to all four different temperatures (marked by open symbols). The color codes for T_{train} which changes smoothly from red ($T = 0.64$, high temperature) to blue ($T = 0.44$, low temperature).

dynamics relationships learnt at a given temperature to make predictions at a different temperature. The results are actually quite remarkable as transferability is generically quite good for all models. This shows that these relationships only evolve quite smoothly across the range of temperatures investigated here. In particular, the models trained at $1/T = 2.0$ ($\tau_\alpha = 210$) perform nearly as good in predicting propensity at $1/T = 2.25$ ($\tau_\alpha = 4100$) as the models trained directly on $1/T = 2.25$. The SE(3) method seems to be particularly suited to transfer to lower temperatures, opening the possibility to study structural relaxation at much lower temperatures.

Task C. Finally, we investigate the performance of the models to predict the correct extent of dynamic heterogeneity. A time-dependent scalar that quantifies heterogeneities is the dynamic susceptibility

$$\chi_4(t) = N (\langle \bar{C}_{\mathcal{R}}^2(t) \rangle - \langle \bar{C}_{\mathcal{R}}(t) \rangle^2) \quad (2)$$

calculated from the system-averaged overlap function $\bar{C}_{\mathcal{R}}(t) = (1/N) \sum_{i \in N} \Theta(0.3 - \mathcal{R}_i(t))$. Here, N is the number of particles in the system, and $\Theta(x)$ the Heaviside function. This definition separates particles into active ($\mathcal{R}_i > 0.3$) and frozen ($\mathcal{R}_i \leq 0.3$). The threshold value of 0.3 is a common choice [121, 122] and corresponds to values slightly larger than the plateau in the mean-squared displacement [121], implying that particles identified as active have typically left their initial cages. In Fig. 6, we compare the results of the predictions to the ground truth MD simulations. Despite the overall good performance in the Pearson correlation, we now observe strong differences between the various techniques. For example, the improvement in performance of SE(3) compared to BOTAN shown in Fig. 4 can be connected to their different learning of the correct dynamic heterogeneity.

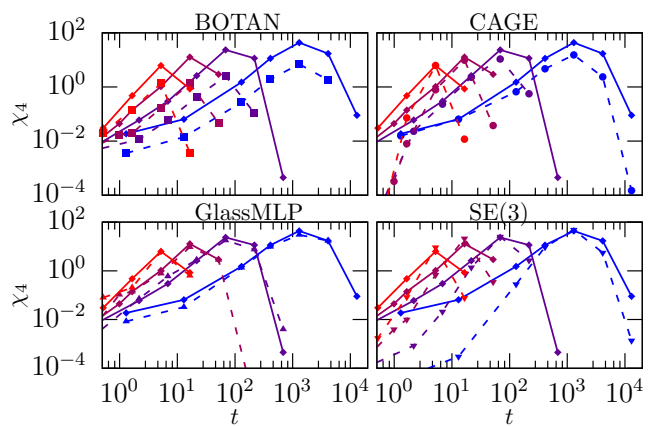


FIG. 6. **Benchmark task (C)** Dynamic susceptibility $\chi_4(t)$ for different temperatures ($T = 0.44, 0.5, 0.56, 0.64$), as predicted from ML techniques (dashed lines), compared to the ground truth (full lines) in the KA system (3D). The color code is as in Fig. 5.

In a recent attention-based GNN extension, this heterogeneity has been explicitly targeted during the training procedure to improve the performance of the deep network [123]. The best overall performance in predicting $\chi_4(t)$ is achieved by GlassMLP, which was specifically constructed to learn and predict dynamic heterogeneity [30]. This analysis shows that the Pearson correlation is not entirely sufficient to quantify the performance of a model. Additional dynamical observables, such as the dynamic susceptibility $\chi_4(t)$, should be investigated to better characterize the ability of models to realistically describe structural relaxation in glass-forming liquids.

In the SM [54] we provide further benchmarking, by investigating: (a) the coefficient of determination R^2 , another popular measure to study the performance of ML models; (b) the probability distributions of the predicted propensities, which correspond to a second contribution to dynamic heterogeneity; (c,d) the scatter plots and the snapshots directly comparing the true and predicted propensities; (e) the cross-correlations between the different structural descriptors and the models investigated in this section. We also provide additional information on (f) learning curves, (g) transferability to lower temperatures $T = 0.4$, and (h) bond-breaking propensity.

Another conclusion of additional benchmarks, which goes beyond the scope of this roadmap, is that correlation coefficients display some system dependence [73]. Hard sphere glasses and systems with strong icosahedral order, for example, show systematically higher correlation coefficients between structural and dynamical descriptors than Kob-Andersen mixtures [24, 51]. The same statement holds true for different dynamical descriptors and coarse-grained quantities. We therefore strongly encourage the use of identical datasets and labels (i.e., dynamical descriptors) to enable comparability.

Among possible extensions of GlassBench, it would

be particularly worth including a diverse set of models of glass-formers: fragile molecular glass-formers, strong network-forming glasses and potentially metallic glasses. The first step in creating a new dataset is the sampling of independent configurations at a given temperature or density. This step could leverage enhanced sampling techniques such as parallel tempering [124] or swap Monte Carlo [125]. Subsequently, for each structure, molecular dynamics or ab-initio dynamics simulations will have to be performed to study the structural relaxation of the systems and calculate the isoconfigurational average [87].

It would also be interesting to broaden the tasks. For example, other ML models, such as the softness derived from SVC, target predicting local energy barriers for short-time rearrangements instead of long-time structural relaxation [19, 85] (see Sec. IV). For this class of ML techniques, it would be preferable that the descriptor can effectively separate the particles by their probability of rearrangement, $P_R(S)$. To encourage further development in this area, it would be desirable to perform a similar benchmarking and investigate whether modern ML models can outperform the state of the art.

Similarly, there is much interest in understanding plastic events and failure of glassy materials under external load. Contrarily to equilibrium structural relaxation, this protocol takes place far from equilibrium, but displays similar characteristics. Recently, this field has been reviewed and benchmarked in a collaborative publication [11]. The focus, however, was not yet on advanced ML techniques, which further demonstrates the very quick development of this field. It would be interesting to analyse whether the techniques presented in this roadmap can help investigating glass deformation under shear [21, 126–128].

VI. SUMMARY AND FURTHER DIRECTIONS

One of the main challenges in understanding glass-forming liquids is connecting their structural properties to emergent relaxation dynamics and plastic deformation. Since the amorphous structure lacks apparent long-range order, machine learning techniques are potentially useful to extract important information from simulations or experimental measurements and eventually generate new data. We therefore anticipate that machine learning will have a decisive influence on the field of glass physics in the upcoming years.

We have given both an overview of recent achievements in using machine learning to advance our fundamental understanding of glass physics, and discussed important research questions for the future. We identified three main fields in which machine learning could impact research on glass physics: (i) automatic feature detection to better understand locally favored amorphous structures, (ii) forecasting of future microscopic dynamics and investigation of dynamical heterogeneities, and (iii) the

construction of data-driven effective models of glassy dynamics. For each of these fields we have proposed concrete methodology that could be used to answer some of the identified research questions. Finally, we have provided a dataset and benchmarking to provide a common ground to compare ML methods for glassy dynamics and enable reproducibility. This work will hopefully provide inspiration and guidelines to researchers on how to further develop the field with the ultimate goal to better understand universal properties of glass-forming liquids and, eventually, the glass transition itself.

The number of potential applications of ML techniques for studies of glass-forming liquids and glasses goes of course much further than what we covered in this article, focusing on fundamental aspects of glassy dynamics. This includes investigation of specific material properties [69, 70, 129, 130] or material discovery [67, 131, 132], machine learning force fields [133], and particle identification in experimental data [134, 135]. Furthermore, the connection between ML and glassy physics has also gone in the reverse direction, by borrowing methods developed for the study of disordered systems to analyze central theoretical ML problems. In fact, the connection between the rough energy landscape of amorphous materials and optimization defined by loss functions with many local minima has been used to better understand and optimize learning of neural networks [136–141].

We anticipate that modern ML frameworks will continue to impact glass research and lead to the development of new major directions in the field. We close this roadmap by discussing exciting new concepts that have the potential to play an important role in future research.

A. Attention and transformers

An important advance in recent ML architectures is the “attention mechanism” [142]. Using it in ML methods for glassy dynamics has a lot of potential. The fundamental concept behind attention is to assign a learnable level of importance to specific parts of the input or intermediate representation. This could be distinct words in sentences (see, e.g., ChatGPT), amorphous structures’ specific features, channels in a deep representation or neighboring atoms in a graph representation. Broadly speaking, this can be achieved by making learned weights themselves dependent on the input.

Salient examples of successful applications of attention are AlphaFold v2 and RoseTTAFold [143, 144], which both use a rotation-equivariant attention-based transformer to predict the three-dimensional structure of proteins. This architecture differs from the graph network architectures discussed earlier by the dense character of its computation mechanism: all atoms in the input can exchange information with all other atoms (with a learned modulation as a function of the distance), allowing for a more flexible computation. Variants of these architectures tend to obtain results which are competitive

with sparse graph networks on benchmark tasks [145]. Recently, a network with a self-attention mechanism was developed for predicting glassy dynamics [123], showing that the curse of overfitting can be avoided. The broader concept of input-dependent weights was also used in a similar context to learn dynamic heterogeneities over a wide range of temperatures at which attention must be paid to temperature-dependent length scales [89]. Both of the approaches are, however, rather direct applications of the idea of positional encoding for learning attention weights, and more complex networks, such as full-fledged transformers [142] are expected to be used in future research.

Going in the opposite direction from transformers, attention mechanisms can also be used to make models more interpretable. For example, they can help identify the task-relevant sectors in the input data [146]. Additionally, attention can also be incorporated into other architectures, for example in combination with a temporal encoding in time-series forecasting [147], to identify relevant parts of past trajectories.

B. Self-supervised, semi-supervised and reinforcement learning

So far, we have mainly focused on “traditional” unsupervised and supervised learning techniques since they are better established in the field. However, there are several other learning paradigms which are potentially useful for future projects, some of which have started to be used in glass research.

To learn the connection between structural order and structural relaxation in a way that reconciles the unsupervised and supervised approaches, a possibility is to use semi-supervised learning [148]. Concretely, the idea is to perform self-supervised learning using only unlabeled input configurations by designing a pretext task, such as reinserting a particle which has been artificially removed from a configuration, denoising particles’ positions, or predicting local quantities (such as the potential energy of each particle or its distance to its quenched position) [91]. Once a representation has been learned to perform this mock task, one can fine-tune only a handful of parameters to correlate the learned representation to the relevant dynamical variable. Such self-supervised pre-training has proven effective in increasing performance for various downstream tasks that are similar in spirit to glassy dynamics prediction, such as molecular properties [149], crystalline material properties [150] or organic semiconductors optoelectronic properties prediction [151]. In the application to glasses, the key element of this approach is to build most of the network without looking at labels, so the output of such a method may be more acceptable as a bona fide structural descriptor, as opposed to heavy networks relying purely on supervised learning. Additionally, it requires fewer labeled data which becomes very important at low tem-

peratures at which sampling becomes difficult. The described methodology could therefore also be used to improve transferability of pre-trained models.

Although reinforcement learning is a well established tool in the field of machine learning [152] only very recently it has found applications in physics for improved sampling [153, 154] or structure optimization [155, 156]. The general idea behind reinforcement learning is to learn taking specific actions when reaching certain states. The goal is to find the policy of actions leading to optimal results, as quantified by a reward function. Applying this approach to the example of searching ground states in spin glasses, the state would be the observed structure, action would be a spin-flip and reward the energy change after several spin flips [154]. Along these lines, Bojesen [157] has formulated the Metropolis-Hastings algorithm in a reinforcement learning setting suitable for simulations of spin systems. Recent work by Galliano *et al.* has extended this approach to learn novel Monte Carlo moves that accelerate sampling of supercooled liquids [158], establishing connections with related adaptive Monte Carlo methods [159, 160]. One of the goals of this line of research is to improve and generalize the swap Monte Carlo algorithm, which demonstrated an impressive performance for specifically adapted glass models [125, 161] and has led to a series of exciting new insights into supercooled liquids [106, 117]. Devising general-purpose enhanced sampling algorithms to simulate glassy systems represents an exciting challenge for future research [162].

C. Generative models

Another promising emerging direction for applying ML techniques to problems relevant to fundamental aspects of glass physics is the use of generative models (GM). One of the key problems in theoretical studies of glassy dynamics is that of sampling. In fact, sampling efficiently configurations x_i from a Gibbs distribution of the form $P(x_i) = \exp[-\beta U(x_i)]/Z$, where $\beta^{-1} = k_B T$ is the target inverse temperature, and $U(x_i)$ is the known potential energy is very challenging for systems exhibiting glassy dynamics. Different from generative modelling of images, in which one estimates an unknown probability distribution from data, here the target distribution is therefore known from the start and sampling from it is the challenge.

A first line of research was proposed independently by Noé *et al.* under the name of Boltzmann generators [163] and by Wu *et al.* [164] using variational autoregressive networks. The idea is to consider a much simpler distribution $\hat{P}(z)$, such that one can easily sample z_i in a single shot. This can, for example, be a Gaussian model, an autoregressive model, or a Gibbs distribution at very high temperatures at which we can sample efficiently. After learning, independent samples x can then be generated using the invertible map $x_i = f(z_i)$ of the model. By

computing the biased distribution $P_{\text{GM}}(x)$ of the generative model enables unbiasing x_i in a last post-processing step [163]. This ensures that we are precisely sampling from $P(x)$.

In practice the method is only efficient if the weights are almost uniformly distributed. This requires that the generative distribution is as close as possible to the target distribution. To this aim, the machine is trained to minimize the Kullback-Leibler (KL) divergence between the generative and target distributions. Because the KL divergence between two distributions is not symmetric, two choices can be made: (i) Maximum likelihood training, in which one minimizes $D_{KL}(P||P_{\text{GM}})$. This approach has the advantage that P_{GM} has to cover well all the support of $P(x)$. However, it also requires existing samples from $P(x)$, which renders training impractical: we want to train a machine to sample from the target, but for this we need to be able to sample from the target. (ii) Variational or energy-based training, where one minimizes $D_{KL}(P_{\text{GM}}||P)$, which corresponds to the free energy of the generative model. While this choice does not require sampling from P_{GM} it has the important drawback that the generative model might only cover part of the support of the target, which is known as mode collapse.

To deal with these problems, several architectures and training strategies have been proposed. To sample a two-state protein model Noé *et al.* [163] focused on a normalizing flow architecture to represent the map $f(z)$, and a training strategy based on mixing the variational approach with maximum likelihood, for which experimental structures and short molecular dynamics simulations were used. In subsequent work, the group introduced equivariant flows to implement physical symmetries [165], and used a higher-temperature Boltzmann distribution as a prior [166, 167] (see also Refs. [168, 169]). To sample crystalline structures, it was recently proposed to generate displacements from a reference lattice structure instead of absolute particle positions [170]. Gabrié *et al.* [171] introduced a more efficient training strategy based on mixing standard Monte Carlo moves with moves proposed by the generative model. Applying Boltzmann generators to sample supercooled liquids yields performances in the same magnitude as previously known enhanced sampling techniques [172].

Several studies have focused on other models in which sampling is challenging, e.g. spin-glasses, hard optimization problems and lattice field theories [154, 164, 173–185]. It remains unclear how these methods compare to standard ones, or whether the efficiency is universal or

model dependent [186–189].

Finally, it could be worth combining the ML models discussed in Sec. III, which can precisely predict future dynamics, with generative models. The former can be used to detect active regions or even particles which are likely to rearrange, while the latter can subsequently propose new configurations based on local rearrangements.

DATA AVAILABILITY

The dataset “GlassBench” and python scripts used to create the benchmarks presented in Sec. V are publicly available and can be download on zenodo (<https://doi.org/10.5281/zenodo.10118191>).

ACKNOWLEDGMENTS

This roadmap paper originates from discussions and interactions at the AISSAI (AI for science, science for AI) workshop on “Machine Learning Glasses” held in November 2022 in Paris (see [190] for details on the workshop and YouTube for recorded sessions). This workshop has been organized by Giulio Biroli, Ludovic Berthier and Gerhard Jung. We thank all participants for their attendance, discussions and feedback, in particular A. Banerjee, L. Janssen, M. Ruiz Garcia, S. Patinet, C. Scalliet, D. Richard, J. Rottler, O. Dauchot and O. Kukhareenko for their valuable contributions. FSP is supported by a public grant overseen by the French National Research Agency (ANR) through the program UDOPIA, project funded by the ANR-20-THIA-0013-01. FSP was granted access to the HPC resources of IDRIS under the allocation 2022-AD011014066 made by GENCI. HS acknowledges computational resources provided by “Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures (JHPCN)” and “High Performance Computing Infrastructure (HPCI)” in Japan (Project ID: jh230064). AJL is supported by the Simons Foundation via the Investigator Award #327939. In addition, AJL thanks CCB at the Flatiron Institute and the Isaac Newton Institute for Mathematical Sciences under the program “New Statistical Physics in Living Matter” (EPSRC grant EP/R014601/1), for support and hospitality. This work was supported by a grant from the Simons Foundation (#454933, LB, #454935 GB). GB acknowledges funding from the French government under management of Agence Nationale de la Recherche as part of the “Investissements d’avenir” program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute).

[1] M. D. Ediger, C. A. Angell, and S. R. Nagel, Supercooled liquids and glasses, *The Journal of Physical Chemistry* **100**, 13200 (1996).

[2] L. Berthier and M. D. Ediger, Facets of glass physics, *Physics Today* **69**, 40 (2016).

[3] H. Li and Y. Zheng, Recent advances in bulk metallic glasses for biomedical applications, *Acta biomaterialia*

- 36**, 1 (2016).
- [4] M. D. Ediger, Perspective: Highly stable vapor-deposited glasses, *The Journal of Chemical Physics* **147**, 210901 (2017).
 - [5] A. Cavagna, Supercooled liquids for pedestrians, *Physics Reports* **476**, 51 (2009).
 - [6] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature* **521**, 436 (2015).
 - [7] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions, *Journal of big Data* **8**, 1 (2021).
 - [8] C. P. Royall and S. R. Williams, The role of local structure in dynamical arrest, *Physics Reports* **560**, 1 (2015), the role of local structure in dynamical arrest.
 - [9] H. Tanaka, H. Tong, R. Shi, and J. Russo, Revealing key structural features hidden in liquids and glasses, *Nature Reviews Physics* **1**, 333 (2019).
 - [10] S. Marín-Aguilar, H. H. Wensink, G. Foffi, and F. Smallenburg, Tetrahedrality dictates dynamics in hard sphere mixtures, *Phys. Rev. Lett.* **124**, 208005 (2020).
 - [11] 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. Materials* **4**, 113609 (2020).
 - [12] E. Boattini, S. Marín-Aguilar, S. Mitra, G. Foffi, F. Smallenburg, and L. Filion, Autonomously revealing hidden local structures in supercooled liquids, *Nature Communications* **11**, 5479 (2020).
 - [13] J. Paret, R. L. Jack, and D. Coslovich, Assessing the structural heterogeneity of supercooled liquids through community inference, *The Journal of Chemical Physics* **152**, 144502 (2020).
 - [14] N. Oyama, S. Koyama, and T. Kawasaki, What do deep neural networks find in disordered structures of glasses?, *Frontiers in Physics* **10**, 1320 (2023).
 - [15] 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).
 - [16] A. Widmer-Cooper and P. Harrowell, Predicting the long-time dynamic heterogeneity in a supercooled liquid? on the basis of short-time heterogeneities, *Physical review letters* **96**, 185701 (2006).
 - [17] A. Widmer-Cooper, H. Perry, P. Harrowell, and D. R. Reichman, Irreversible reorganization in a supercooled liquid originates from localized soft modes, *Nature Physics* **4**, 711 (2008).
 - [18] L. Berthier and G. Biroli, Theoretical perspective on the glass transition and amorphous materials, *Rev. Mod. Phys.* **83**, 587 (2011).
 - [19] S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, A structural approach to relaxation in glassy liquids, *Nature Physics* **12**, 469 (2016).
 - [20] 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** (2015).
 - [21] 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, *Nature Physics* **16**, 448 (2020).
 - [22] 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).
 - [23] E. Boattini, F. Smallenburg, and L. Filion, Averaging local structure to predict the dynamic propensity in supercooled liquids, *Phys. Rev. Lett.* **127**, 088007 (2021).
 - [24] R. M. Alkemade, E. Boattini, L. Filion, and F. Smallenburg, Comparing machine learning techniques for predicting glassy dynamics, *The Journal of Chemical Physics* **156**, 204503 (2022).
 - [25] H. Shiba, M. Hanai, T. Suzumura, and T. Shimokawabe, BOTAN: BOnd TArgeting Network for prediction of slow glassy dynamics by machine learning relative motion, *The Journal of Chemical Physics* **158**, 084503 (2023).
 - [26] R. M. Alkemade, F. Smallenburg, and L. Filion, Improving the prediction of glassy dynamics by pinpointing the local cage, *The Journal of Chemical Physics* **158**, 134512 (2023).
 - [27] S. Ciarella, M. Chiappini, E. Boattini, M. Dijkstra, and L. M. C. Janssen, Dynamics of supercooled liquids from static averaged quantities using machine learning, *Machine Learning: Science and Technology* **4**, 025010 (2023).
 - [28] F. S. Pezzicoli, G. Charpiat, and F. P. Landes, SE(3)-equivariant graph neural networks for learning glassy liquids representations, arXiv preprint arXiv:2211.03226 (2022).
 - [29] M. Ruiz-Garcia, C. M. Barriuso G., L. C. Alexander, D. G. A. L. Aarts, L. M. Ghiringhelli, and C. Valeriani, Discovering dynamic laws from observations: The case of self-propelled, interacting colloids, *Phys. Rev. E* **109**, 064611 (2024).
 - [30] 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).
 - [31] 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).
 - [32] 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).
 - [33] See, e.g., <https://paperswithcode.com/sota> for benchmarks in the fields of computer vision, natural language processing, time series analysis and much more.
 - [34] G. Tarjus, D. Kivelson, and P. Viot, The viscous slowing down of supercooled liquids as a temperature-controlled super-Arrhenius activated process: A description in terms of frustration-limited domains, *J. Phys.: Condens. Matter* **12**, 6497 (2000).
 - [35] M. Tanemura, Y. Hiwatari, H. Matsuda, T. Ogawa, N. Ogita, and A. Ueda, Geometrical Analysis of Crystallization of the Soft-Core Model, *Prog Theor Phys* **58**, 1079 (1977).

- [36] A. Malins, S. R. Williams, J. Eggers, and C. P. Royall, Identification of structure in condensed matter with the topological cluster classification, *J. Chem. Phys.* **139**, 234506 (2013).
- [37] J. D. Honeycutt and H. C. Andersen, Molecular dynamics study of melting and freezing of small Lennard-Jones clusters, *J. Phys. Chem.* **91**, 4950 (1987).
- [38] E. A. Lazar, J. Han, and D. J. Srolovitz, A Topological Framework for Local Structure Analysis in Condensed Matter, *Proc. Natl. Acad. Sci.* **112**, E5769 (2015).
- [39] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Bond-orientational order in liquids and glasses, *Phys. Rev. B* **28**, 784 (1983).
- [40] P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to Machine Learning for physicists, *Physics Reports* **810**, 1 (2019).
- [41] B. Cheng, R.-R. Griffiths, S. Wengert, C. Kunkel, T. Stenczel, B. Zhu, V. L. Deringer, N. Bernstein, J. T. Margraf, K. Reuter, and G. Csányi, Mapping Materials and Molecules, *Acc. Chem. Res.* **53**, 1981 (2020).
- [42] A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé, and A. Laio, Unsupervised Learning Methods for Molecular Simulation Data, *Chem. Rev.* **121**, 9722 (2021).
- [43] W. Hu, R. R. P. Singh, and R. T. Scalettar, Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination, *Phys. Rev. E* **95**, 062122 (2017).
- [44] J. F. Rodriguez-Nieva and M. S. Scheurer, Identifying topological order through unsupervised machine learning, *Nat. Phys.* **15**, 790 (2019).
- [45] T. Mendes-Santos, X. Turkeshi, M. Dalmonte, and A. Rodriguez, Unsupervised Learning Universal Critical Behavior via the Intrinsic Dimension, *Phys. Rev. X* **11**, 011040 (2021).
- [46] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, *Phys. Rev. B* **87**, 184115 (2013).
- [47] B. Parsaeifard, D. S. De, A. S. Christensen, F. A. Faber, E. Kocer, S. De, J. Behler, A. von Lilienfeld, and S. Goedecker, An assessment of the structural resolution of various fingerprints commonly used in machine learning, *Mach. Learn.: Sci. Technol.* **2**, 015018 (2021), arXiv:2008.03189.
- [48] B. Midtvedt, J. Pineda, F. Skärberg, E. Olsén, H. Bachimanchi, E. Wesén, E. K. Esbjörner, E. Selander, F. Höök, D. Midtvedt, *et al.*, Single-shot self-supervised object detection in microscopy, *Nature Communications* **13**, 7492 (2022).
- [49] M. A. Caro, V. L. Deringer, J. Koskinen, T. Laurila, and G. Csányi, Growth Mechanism and Origin of High sp^3 in Tetrahedral Amorphous Carbon, *Phys. Rev. Lett.* **120**, 166101 (2018).
- [50] B. Monserrat, J. G. Brandenburg, E. A. Engel, and B. Cheng, Liquid water contains the building blocks of diverse ice phases, *Nat Commun* **11**, 5757 (2020).
- [51] D. Coslovich, R. L. Jack, and J. Paret, Dimensionality reduction of local structure in glassy binary mixtures, *J. Chem. Phys.* **157**, 204503 (2022).
- [52] A. Banerjee, H.-P. Hsu, K. Kremer, and O. Kukhareenko, Data-driven identification and analysis of the glass transition in polymer melts, *ACS Macro Letters* **12**, 679 (2023).
- [53] A. Banerjee, A. Iscen, K. Kremer, and O. Kukhareenko, Determining glass transition in all-atom acrylic polymeric melt simulations using machine learning, *The Journal of Chemical Physics* **159**, 074108 (2023).
- [54] See supplemental material (SM) [URL will be inserted by the publisher] for technical details and additional benchmarking. The SM contains Refs. [106, 161, 191–195].
- [55] A. Offei-Danso, A. Hassanali, and A. Rodriguez, High-Dimensional Fluctuations in Liquid Water: Combining Chemical Intuition with Unsupervised Learning, *J. Chem. Theory Comput.* **18**, 3136 (2022).
- [56] P. Campadelli, E. Casiraghi, C. Ceruti, and A. Rozza, Intrinsic Dimension Estimation: Relevant Techniques and a Benchmark Framework, *Math. Probl. Eng.* **2015**, e759567 (2015).
- [57] Note that in the presence of agnostic, high-dimensional structural descriptors, such as the smooth atomic overlap parameters [46], interpretation always occurs *a posteriori*, by searching for correlations between some of the reduced structural and physically-motivated structural measures [41].
- [58] B. Parsaeifard and S. Goedecker, Manifolds of quasi-constant SOAP and ACSF fingerprints and the resulting failure to machine learn four-body interactions, *J. Chem. Phys.* **156**, 034302 (2022).
- [59] J. P. Darby, J. R. Kermode, and G. Csányi, Compressing local atomic neighbourhood descriptors, *npj Comput Mater* **8**, 1 (2022).
- [60] J. P. Darby, D. P. Kovács, I. Batatia, M. A. Caro, G. L. W. Hart, C. Ortner, and G. Csányi, Tensor-reduced atomic density representations, *Phys. Rev. Lett.* **131**, 028001 (2023).
- [61] D. Coslovich, M. Ozawa, and L. Berthier, Local order and crystallization of dense polydisperse hard spheres, *Journal of Physics: Condensed Matter* **30**, 144004 (2018).
- [62] H. Tong and H. Tanaka, Emerging exotic compositional order on approaching low-temperature equilibrium glasses, *Nature Communications* **14**, 4614 (2023).
- [63] S. R. Elliott, Medium-range structural order in covalent amorphous solids, *Nature* **354**, 445–452 (1991).
- [64] H. Sheng, W. Luo, F. Alamgir, J. Bai, and E. Ma, Atomic packing and short-to-medium-range order in metallic glasses, *Nature* **439**, 419 (2006).
- [65] J. M. Montes de Oca, F. Sciortino, and G. A. Appignanesi, A structural indicator for water built upon potential energy considerations, *The Journal of Chemical Physics* **152**, 244503 (2020).
- [66] C. Faccio, M. Benzi, L. Zanetti-Polzi, and I. Daidone, Low- and high-density forms of liquid water revealed by a new medium-range order descriptor, *Journal of Molecular Liquids* **355**, 118922 (2022).
- [67] J. C. Mauro, A. Tandia, K. D. Vargheese, Y. Z. Mauro, and M. M. Smedskjaer, Accelerating the design of functional glasses through modeling, *Chemistry of Materials* **28**, 4267 (2016).
- [68] D. R. Cassar, S. M. Mastelini, T. Botari, E. Alcobaça, A. C. P. L. F. de Carvalho, and E. D. Zanotto, Predicting and interpreting oxide glass properties by machine learning using large datasets, *Ceramics International* **47**, 23958–23972 (2021).
- [69] M. L. Bødker, M. Bauchy, T. Du, J. C. Mauro, and M. M. Smedskjaer, Predicting glass structure by

- physics-informed machine learning, *npj Computational Materials* **8**, 192 (2022).
- [70] Ravinder, V. Venugopal, S. Bishnoi, S. Singh, M. Zaki, H. S. Grover, M. Bauchy, M. Agarwal, and N. A. Krishnan, Artificial intelligence and machine learning in glass science and technology: 21 challenges for the 21st century, *International journal of applied glass science* **12**, 277 (2021).
- [71] 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).
- [72] 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).
- [73] 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).
- [74] 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).
- [75] Note that in Ref. [20] the term Support Vector Machine was used, but here we use the term Support Vector Classifier to stress the fact that the procedure corresponded to a classification and not a regression.
- [76] S. S. Schoenholz, E. D. Cubuk, E. Kaxiras, and A. J. Liu, Relationship between local structure and relaxation in out-of-equilibrium glassy systems, *Proceedings of the National Academy of Sciences* **114**, 263 (2017).
- [77] D. M. Sussman, S. S. Schoenholz, E. D. Cubuk, and A. J. Liu, Disconnecting structure and dynamics in glassy thin films, *Proceedings of the National Academy of Sciences* **114**, 10601 (2017).
- [78] E. D. Cubuk, R. J. S. Ivancic, S. S. Schoenholz, D. J. Strickland, A. Basu, Z. S. Davidson, J. Fontaine, J. L. Hor, Y. R. Huang, Y. Jiang, N. C. Keim, K. D. Koshigan, J. A. Lefever, T. Liu, X. G. Ma, D. J. Magagnosc, E. Morrow, C. P. Ortiz, J. M. Rieser, A. Shavit, T. Still, Y. Xu, Y. Zhang, K. N. Nordstrom, P. E. Arratia, R. W. Carpick, D. J. Durian, Z. Fakhraai, D. J. Jerolmack, D. Lee, J. Li, R. Riggleman, K. T. Turner, A. G. Yodh, D. S. Gianola, and A. J. Liu, Structure-property relationships from universal signatures of plasticity in disordered solids, *Science* **358**, 1033 (2017).
- [79] M. Harrington, A. J. Liu, and D. J. Durian, Machine learning characterization of structural defects in amorphous packings of dimers and ellipses, *Phys. Rev. E* **99** (2019).
- [80] X. Ma, Z. S. Davidson, T. Still, R. J. S. Ivancic, S. S. Schoenholz, A. J. Liu, and A. G. Yodh, Heterogeneous activation, local structure, and softness in supercooled colloidal liquids, *Phys. Rev. Lett.* **122** (2019).
- [81] E. D. Cubuk, A. J. Liu, E. Kaxiras, and S. S. Schoenholz, Unifying framework for strong and fragile liquids via machine learning: a study of liquid silica, arXiv preprint arXiv:2008.09681 (2020).
- [82] S. A. Ridout, J. W. Rocks, and A. J. Liu, Correlation of plastic events with local structure in jammed packings across spatial dimensions, *Proceedings of the National Academy of Sciences* **119** (2022).
- [83] S. A. Tah, I. Ridout and A. J. Liu, Fragility in glassy liquids: a structural approach based on machine learning, *J. Chem. Phys.* **157** (2022).
- [84] H. Liu, M. M. Smedskjaer, and M. Bauchy, Deciphering a structural signature of glass dynamics by machine learning, *Phys. Rev. B* **106**, 214206 (2022).
- [85] G. Zhang, S. A. Ridout, and A. J. Liu, Interplay of rearrangements, strain, and local structure during avalanche propagation, *Phys. Rev. X* **11** (2021).
- [86] H. Xiao, G. Zhang, E. Yang, R. J. Ivancic, S. A. Ridout, R. Riggleman, D. J. Durian, and A. J. Liu, Machine learning-informed structuro-elastoplasticity predicts ductility of disordered solids, arXiv preprint arXiv:2303.12486 (2023).
- [87] A. Widmer-Cooper, P. Harrowell, and H. Fynewever, How reproducible are dynamic heterogeneities in a supercooled liquid?, *Phys. Rev. Lett.* **93**, 135701 (2004).
- [88] L. Berthier and R. L. Jack, Structure and dynamics of glass formers: Predictability at large length scales, *Phys. Rev. E* **76**, 041509 (2007).
- [89] G. Jung, G. Biroli, and L. Berthier, Dynamic heterogeneity at the experimental glass transition predicted by transferable machine learning, *Phys. Rev. B* **109**, 064205 (2024).
- [90] L. Berthier, G. Biroli, J.-P. Bouchaud, L. Cipelletti, and W. Saarloos, *Dynamical Heterogeneities in Glasses, Colloids, and Granular Media* (Oxford University Press, 2011) pp. 1–464.
- [91] S. Gidaris, P. Singh, and N. Komodakis, Unsupervised representation learning by predicting image rotations, arXiv preprint arXiv:1803.07728 (2018).
- [92] C. Toninelli, M. Wyart, L. Berthier, G. Biroli, and J.-P. Bouchaud, Dynamical susceptibility of glass formers: Contrasting the predictions of theoretical scenarios, *Phys. Rev. E* **71**, 041505 (2005).
- [93] D. P. Kingma, M. Welling, *et al.*, An introduction to variational autoencoders, *Foundations and Trends® in Machine Learning* **12**, 307 (2019).
- [94] Q. Wang and L. Zhang, Inverse design of glass structure with deep graph neural networks, *Nature communications* **12**, 5359 (2021).
- [95] S. Kivelson and S. Kivelson, Understanding complexity, *Nature Physics* **14**, 426–427 (2018).
- [96] Q. Wang, J. Ding, L. Zhang, E. Podryabinkin, A. Shapeev, and E. Ma, Predicting the propensity for thermally activated β events in metallic glasses via interpretable machine learning, *npj Computational Materials* **6**, 194 (2020).
- [97] S. Miao, M. Liu, and P. Li, Interpretable and generalizable graph learning via stochastic attention mechanism, in *International Conference on Machine Learning* (PMLR, 2022) pp. 15524–15543.
- [98] E. Duede, Deep learning opacity in scientific discovery, *Philosophy of Science* **90**, 1089–1099 (2023).
- [99] A. Glielmo, C. Zeni, B. Cheng, G. Csányi, and A. Laio, Ranking the information content of distance measures, *PNAS Nexus* **1**, pgac039 (2022).
- [100] A. Sharma, C. Liu, and M. Ozawa, Selecting relevant structural features for glassy dynamics by information imbalance, arXiv preprint arXiv:2408.12705 10.48550/arXiv.2408.12705 (2024), arXiv:2408.12705 [cond-mat].
- [101] L. Berthier, E. Flenner, and G. Szamel, Glassy dynamics in dense systems of active particles, *The Journal of Chemical Physics* **150**, 200901 (2019).
- [102] G. Janzen and L. M. C. Janssen, Rejuvenation and memory effects in active glasses induced by thermal and

- active cycling, *Phys. Rev. Res.* **6**, 023257 (2024).
- [103] G. Janzen, X. L. Smeets, V. E. Debets, C. Luo, C. Storm, L. M. Janssen, and S. Ciarella, Dead or alive: Distinguishing active from passive particles using supervised learning (a), *Europhysics Letters* **143**, 17004 (2023).
- [104] G. Janzen, C. Smit, S. Visbeek, V. E. Debets, C. Luo, C. Storm, S. Ciarella, and L. M. C. Janssen, Classifying the age of a glass based on structural properties: A machine learning approach, *Phys. Rev. Mater.* **8**, 025602 (2024).
- [105] C. Scalliet, B. Guiselin, and L. Berthier, Excess wings and asymmetric relaxation spectra in a facilitated trap model, *The Journal of Chemical Physics* **155**, 064505 (2021).
- [106] B. Guiselin, C. Scalliet, and L. Berthier, Microscopic origin of excess wings in relaxation spectra of supercooled liquids, *Nature Physics* **18**, 468 (2022).
- [107] A. Nicolas, E. E. Ferrero, K. Martens, and J.-L. Barrat, Deformation and flow of amorphous solids: Insights from elastoplastic models, *Rev. Mod. Phys.* **90**, 045006 (2018).
- [108] M. Ozawa and G. Biroli, Elasticity, facilitation and dynamic heterogeneity in glass-forming liquids, *Phys. Rev. Lett.* **130**, 138201 (2023).
- [109] A. Tahaei, G. Biroli, M. Ozawa, M. Popović, and M. Wyart, Scaling description of dynamical heterogeneity and avalanches of relaxation in glass-forming liquids, *Phys. Rev. X* **13**, 031034 (2023).
- [110] 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).
- [111] R. N. Chacko, F. P. Landes, G. Biroli, O. Dauchot, A. J. Liu, and D. R. Reichman, Elastoplasticity mediates dynamical heterogeneity below the mode coupling temperature, *Phys. Rev. Lett.* **127** (2021).
- [112] C. Monthus and J. P. Bouchaud, Models of traps and glass phenomenology, *J. Phys. A* **29** (1996).
- [113] S. A. Ridout, I. Tah, and A. J. Liu, Building a “trap model” of glassy dynamics from a local structural predictor of rearrangements, *Europhysics Letters* **144**, 47001 (2023).
- [114] S. A. Ridout and A. J. Liu, The dynamics of machine-learned “softness” in supercooled liquids describe dynamical heterogeneity, arXiv preprint arXiv:2406.05868 (2024).
- [115] S. Ciarella, D. Khomenko, L. Berthier, F. C. Mocanu, D. R. Reichman, C. Scalliet, and F. Zamponi, Finding defects in glasses through machine learning, *Nature Communications* **14**, 4229 (2023).
- [116] D. Richard, G. Kapteijns, and E. Lerner, Detecting low-energy quasilocalized excitations in computer glasses, *Phys. Rev. E* **108**, 044124 (2023).
- [117] C. Scalliet, B. Guiselin, and L. Berthier, Thirty milliseconds in the life of a supercooled liquid, *Phys. Rev. X* **12**, 041028 (2022).
- [118] M. D. Ediger, Spatially heterogeneous dynamics in supercooled liquids, *Annual Review of Physical Chemistry* **51**, 99 (2000).
- [119] H. Tong and H. Tanaka, Structural order as a genuine control parameter of dynamics in simple glass formers, *Nature communications* **10**, 5596 (2019).
- [120] Calculating the Pearson correlation for each structure individually and then averaging yields slightly different results, and Pearson correlations appear to be systematically higher. Similarly, calculating Pearson correlation over particles of different type significantly increases the correlation and this should be avoided.
- [121] N. Lačević, F. W. Starr, T. B. Schröder, and S. C. Glotzer, Spatially heterogeneous dynamics investigated via a time-dependent four-point density correlation function, *The Journal of Chemical Physics* **119**, 7372 (2003).
- [122] 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).
- [123] 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, *The Journal of Chemical Physics* **159**, 144504 (2023).
- [124] R. H. Swendsen and J.-S. Wang, Replica monte carlo simulation of spin-glasses, *Physical review letters* **57**, 2607 (1986).
- [125] 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).
- [126] Z. Fan and E. Ma, Predicting orientation-dependent plastic susceptibility from static structure in amorphous solids via deep learning, *Nature communications* **12**, 1506 (2021).
- [127] T. Du, H. Liu, L. Tang, S. S. Sørensen, M. Bauchy, and M. M. Smedskjaer, Predicting fracture propensity in amorphous alumina from its static structure using machine learning, *ACS Nano* **15**, 17705 (2021), pMID: 34723489.
- [128] F. Font-Clos, M. Zanchi, S. Hiemer, S. Bonfanti, R. Guerra, M. Zaiser, and S. Zapperi, Predicting the failure of two-dimensional silica glasses, *Nature communications* **13**, 2820 (2022).
- [129] H. Liu, Z. Fu, K. Yang, X. Xu, and M. Bauchy, Machine learning for glass science and engineering: A review, *Journal of Non-Crystalline Solids* **557**, 119419 (2021).
- [130] D. R. Cassar, GlassNet: a multitask deep neural network for predicting many glass properties, *Ceramics International* , S0272884223025464 (2023).
- [131] A. Tandia, M. C. Onbasli, and J. C. Mauro, Machine learning for glass modeling, *Springer handbook of glass* , 1157 (2019).
- [132] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, Scaling deep learning for materials discovery, *Nature* **624**, 80 (2023).
- [133] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Machine learning force fields, *Chemical Reviews* **121**, 10142 (2021).
- [134] G. Volpe, C. Wählby, L. Tian, M. Hecht, A. Yachimovich, K. Monakhova, L. Waller, I. F. Sbalzarini, C. A. Metzler, M. Xie, *et al.*, Roadmap on deep learning for microscopy, ArXiv (2023).
- [135] B. Midtvedt, J. Pineda, H. Klein Morberg, C. Manzo, and G. Volpe, *DeepTrack2*.
- [136] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, *Physics reports* **810**, 1 (2019).

- [137] M. Gabrié, Mean-field inference methods for neural networks, *Journal of Physics A: Mathematical and Theoretical* **53**, 223002 (2020).
- [138] A. Merchant, L. Metz, S. S. Schoenholz, and E. D. Cubuk, Learn2hop: Learned optimization on rough landscapes, in *International Conference on Machine Learning* (PMLR, 2021) pp. 7643–7653.
- [139] M. Gabrié, S. Ganguli, C. Lucibello, and R. Zecchina, Neural networks: from the perceptron to deep nets, arXiv preprint arXiv:2304.06636 (2023).
- [140] T. Bonnaire, D. Ghio, K. Krishnamurthy, F. Mignacco, A. Yamamura, and G. Biroli, High-dimensional non-convex landscapes and gradient descent dynamics, arXiv preprint arXiv:2308.03754 (2023).
- [141] M. Mézard, Spin glass theory and its new challenge: structured disorder, arXiv preprint arXiv:2309.06947 (2023).
- [142] A. Vaswani, Attention is all you need, *Advances in Neural Information Processing Systems* (2017).
- [143] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, Highly accurate protein structure prediction with alphafold, *Nature* **596**, 583 (2021).
- [144] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, and D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network, *Science* **373**, 871 (2021).
- [145] L. A. Bratholm, W. Gerrard, B. Anderson, S. Bai, S. Choi, L. Dang, P. Hanchar, A. Howard, S. Kim, Z. Kolter, R. Kondor, M. Kornbluth, Y. Lee, Y. Lee, J. P. Mailoa, T. T. Nguyen, M. Popovic, G. Rakocevic, W. Reade, W. Song, L. Stojanovic, E. H. Thiede, N. Tijanic, A. Torrubia, D. Willmott, C. P. Butts, and D. R. Glowacki, A community-powered search of machine learning strategy space to find NMR property prediction models, *PLOS ONE* **16**, e0253612 (2021).
- [146] S. Miao, M. Liu, and P. Li, Interpretable and generalizable graph learning via stochastic attention mechanism, in *Proceedings of the 39th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 162, edited by K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato (PMLR, 2022) pp. 15524–15543.
- [147] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, A dual-stage attention-based recurrent neural network for time series prediction, arXiv preprint arXiv:1704.02971 (2017).
- [148] O. Chapelle, B. Schölkopf, and A. Zien, eds., *Semi-Supervised Learning* (The MIT Press, 2006).
- [149] Y. Rong, Y. Bian, T. Xu, W. Xie, Y. Wei, W. Huang, and J. Huang, Self-supervised graph transformer on large-scale molecular data, *Advances in Neural Information Processing Systems* **33**, 12559 (2020).
- [150] R. Magar, Y. Wang, and A. Barati Farimani, Crystal twins: self-supervised learning for crystalline material property prediction, *npj Computational Materials* **8**, 231 (2022).
- [151] Z. ZHANG, Q. Liu, S. Zhang, C.-Y. Hsieh, L. Shi, and C.-K. Lee, Graph self-supervised learning for optoelectronic properties of organic semiconductors, in *ICML 2022 2nd AI for Science Workshop* (2022).
- [152] L. P. Kaelbling, M. L. Littman, and A. W. Moore, Reinforcement learning: A survey, *Journal of artificial intelligence research* **4**, 237 (1996).
- [153] K. Shin, D. P. Tran, K. Takemura, A. Kitao, K. Terayama, and K. Tsuda, Enhancing biomolecular sampling with reinforcement learning: A tree search molecular dynamics simulation method, *ACS Omega* **4**, 13853 (2019), PMID: 31497702.
- [154] C. Fan, M. Shen, Z. Nussinov, Z. Liu, Y. Sun, and Y.-Y. Liu, Searching for spin glass ground states through deep reinforcement learning, *Nature Communications* **14**, 725 (2023).
- [155] K. Ahuja, W. H. Green, and Y.-P. Li, Learning to optimize molecular geometries using reinforcement learning, *Journal of Chemical Theory and Computation* **17**, 818 (2021), PMID: 33470813.
- [156] V. Bihani, S. Manchanda, S. Sastry, S. Ranu, and N. A. Krishnan, Stridernet: A graph reinforcement learning approach to optimize atomic structures on rough energy landscapes, in *International Conference on Machine Learning* (PMLR, 2023) pp. 2431–2451.
- [157] T. A. Bojesen, Policy-guided monte carlo: Reinforcement-learning markov chain dynamics, *Physical Review E* **98**, 063303 (2018).
- [158] L. Galliano, R. Rende, and D. Coslovich, Policy-guided monte carlo on general state spaces: Application to glass-forming mixtures (2024), arXiv:2407.03275 [cond-mat.soft].
- [159] H. Christiansen, F. Errica, and F. Alesiani, Self-tuning hamiltonian monte carlo for accelerated sampling, *The Journal of Chemical Physics* **159**, 234109 (2023).
- [160] M. Gabrié, G. M. Rotskoff, and E. Vanden-Eijnden, Adaptive monte carlo augmented with normalizing flows, *Proceedings of the National Academy of Sciences* **119**, e2109420119 (2022).
- [161] A. Ninarello, L. Berthier, and D. Coslovich, Models and algorithms for the next generation of glass transition studies, *Phys. Rev. X* **7**, 021039 (2017).
- [162] L. Berthier and D. R. Reichman, Modern computational studies of the glass transition, *Nature Reviews Physics* **5**, 102 (2023).
- [163] F. Noé, S. Olsson, J. Köhler, and H. Wu, Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning, *Science* **365**, eaaw1147 (2019).
- [164] D. Wu, L. Wang, and P. Zhang, Solving statistical mechanics using variational autoregressive networks, *Phys. Rev. Lett.* **122**, 080602 (2019).
- [165] J. Köhler, L. Klein, and F. Noé, Equivariant flows: exact likelihood generative learning for symmetric densities, in *International conference on machine learning* (PMLR, 2020) pp. 5361–5370.

- [166] M. Dibak, L. Klein, A. Krämer, and F. Noé, Temperature steerable flows and boltzmann generators, *Phys. Rev. Res.* **4**, L042005 (2022).
- [167] M. Invernizzi, A. Krämer, C. Clementi, and F. Noé, Skipping the replica exchange ladder with normalizing flows, *The Journal of Physical Chemistry Letters* **13**, 11643 (2022).
- [168] M. Xu, L. Yu, Y. Song, C. Shi, S. Ermon, and J. Tang, Geodiff: A geometric diffusion model for molecular conformation generation, arXiv preprint arXiv:2203.02923 (2022).
- [169] A. Coretti, S. Falkner, P. Geissler, and C. Dellago, Learning mappings between equilibrium states of liquid systems using normalizing flows, arXiv preprint arXiv:2208.10420 (2022).
- [170] S. van Leeuwen, A. P. de Alba Ortíz, and M. Dijkstra, A boltzmann generator for the isobaric-isothermal ensemble (2023), arXiv:2305.08483.
- [171] M. Gabrié, G. M. Rotskoff, and E. Vanden-Eijnden, Adaptive monte carlo augmented with normalizing flows, *Proceedings of the National Academy of Sciences* **119**, e2109420119 (2022).
- [172] G. Jung, G. Biroli, and L. Berthier, Normalizing flows as an enhanced sampling method for atomistic supercooled liquids, *Machine Learning: Science and Technology* **5**, 035053 (2024).
- [173] B. McNaughton, M. Milošević, A. Perali, and S. Pilati, Boosting monte carlo simulations of spin glasses using autoregressive neural networks, *Phys. Rev. E* **101**, 053312 (2020).
- [174] M. Hibat-Allah, E. M. Inack, R. Wiersema, R. G. Melko, and J. Carrasquilla, Variational neural annealing, *Nature Machine Intelligence* **3**, 952 (2021).
- [175] D. Wu, R. Rossi, and G. Carleo, Unbiased monte carlo cluster updates with autoregressive neural networks, *Phys. Rev. Res.* **3**, L042024 (2021).
- [176] E. M. Inack, S. Morawetz, and R. G. Melko, Neural annealing and visualization of autoregressive neural networks in the newman–moore model, *Condensed Matter* **7**, 38 (2022).
- [177] S. Ciarella, J. Trinquier, M. Weigt, and F. Zamponi, Machine-learning-assisted monte carlo fails at sampling computationally hard problems, *Machine Learning: Science and Technology* **4**, 010501 (2023).
- [178] M. J. Schuetz, J. K. Brubaker, Z. Zhu, and H. G. Katzgraber, Graph coloring with physics-inspired graph neural networks, *Phys. Rev. Res.* **4**, 043131 (2022).
- [179] M. J. Schuetz, J. K. Brubaker, and H. G. Katzgraber, Combinatorial optimization with physics-inspired graph neural networks, *Nature Machine Intelligence* **4**, 367 (2022).
- [180] M. S. Albergo, G. Kanwar, and P. E. Shanahan, Flow-based generative models for markov chain monte carlo in lattice field theory, *Phys. Rev. D* **100**, 034515 (2019).
- [181] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racaniere, D. J. Rezende, and P. E. Shanahan, Equivariant flow-based sampling for lattice gauge theory, *Phys. Rev. Lett.* **125**, 121601 (2020).
- [182] P. de Haan, C. Rainone, M. C. Cheng, and R. Bondesan, Scaling up machine learning for quantum field theory with equivariant continuous flows, arXiv preprint arXiv:2110.02673 (2021).
- [183] M. Gerdes, P. de Haan, C. Rainone, R. Bondesan, and M. C. Cheng, Learning lattice quantum field theories with equivariant continuous flows, *SciPost Physics* **15**, 238 (2023).
- [184] D. Luo, G. Carleo, B. K. Clark, and J. Stokes, Gauge equivariant neural networks for quantum lattice gauge theories, *Phys. Rev. Lett.* **127**, 276402 (2021).
- [185] T. Marchand, M. Ozawa, G. Biroli, and S. Mallat, Wavelet conditional renormalization group, arXiv preprint arXiv:2207.04941 (2022).
- [186] M. C. Angelini and F. Ricci-Tersenghi, Modern graph neural networks do worse than classical greedy algorithms in solving combinatorial optimization problems like maximum independent set, *Nature Machine Intelligence* **5**, 29 (2023).
- [187] S. Boettcher, Inability of a graph neural network heuristic to outperform greedy algorithms in solving combinatorial optimization problems, *Nature Machine Intelligence* **5**, 24 (2023).
- [188] S. Boettcher, Deep reinforced learning heuristic tested on spin-glass ground states: The larger picture, *Nature Communications* **14**, 5658 (2023).
- [189] D. Ghio, Y. Dandi, F. Krzakala, and L. Zdeborová, Sampling with flows, diffusion and autoregressive neural networks: A spin-glass perspective, arXiv preprint arXiv:2308.14085 (2023).
- [190] Links to [Workshop](#) and [YouTube channel](#).
- [191] L. Berthier, E. Flenner, C. J. Fullerton, C. Scalliet, and M. Singh, Efficient swap algorithms for molecular dynamics simulations of equilibrium supercooled liquids, *Journal of Statistical Mechanics: Theory and Experiment* **2019**, 064004 (2019).
- [192] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, C. Gulcehre, F. Song, A. Ballard, J. Gilmer, G. Dahl, A. Vaswani, K. Allen, C. Nash, V. Langston, C. Dyer, N. Heess, D. Wierstra, P. Kohli, M. Botvinick, O. Vinyals, Y. Li, and R. Pascanu, Relational inductive biases, deep learning, and graph networks (2018), arXiv:1806.01261.
- [193] J. Paret and D. Coslovich, partycls: A python package for structural clustering, *JOSS* **6**, 3723 (2021).
- [194] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Bond-orientational order in liquids and glasses, *Phys. Rev. B* **28**, 784 (1983).
- [195] 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, *The European Physical Journal E* **41**, 1 (2018).

Supplemental material: Roadmap on machine learning glassy dynamics

In this supplemental material (SM) we provide the technical details necessary to reproduce the data shown in the benchmarking section V of the main manuscript. Additionally, we present further benchmarking.

I. SYSTEMS AND METHODS

A. KA system (3D)

Most benchmarks were performed for the Kob-Andersen binary mixture (KA) in three dimensions [S1]. The basic potential used in KA is the Lennard-Jones potential,

$$V_{\alpha\beta}(r_{ij}) = \begin{cases} 4\epsilon_{\alpha\beta} \left[\left(\frac{\sigma_{\alpha\beta}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{ij}} \right)^6 \right] & r_{ij} < r_{\alpha\beta}^{\text{cut}} \\ 0 & \text{otherwise,} \end{cases}$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between particles i and j , of type α and β , respectively. The KA mixture is characterized by its non-additive interactions, $\epsilon_{11} = 1.0$, $\epsilon_{12} = 1.5$, $\epsilon_{22} = 0.5$ and $\sigma_{11} = 1.0$, $\sigma_{12} = 0.8$, $\sigma_{22} = 0.88$, between the two particles types. The cutoff is $r_{\alpha\beta}^{\text{cut}} = 2.5\sigma_{\alpha\beta}$.

All details on the specific KA mixture and the dataset used in this roadmap are given in Ref. [S2]. The system, for example, includes additional contributions in the potential to make it continuous up to the second derivative. The relaxation dynamics of each structure has been simulated using $N_R = 32$ independent replicas to calculate the isoconfigurational average. All results have been extracted by solely considering type 1 particles. The networks have been trained on $N_S = 400$ structures, with $N_1 = 3277$ type 1 particles each, resulting in a total of $N_d \approx 10^6$ data points. The testset consists of $N_S = 100$ independent structures.

B. KA2D system

The two-dimensional KA2D system is a variation of the above described KA mixture. In addition to the change of dimensionality, KA2D consists of three particle types, instead of two. This change enables to apply the swap Monte Carlo algorithm [S3, S4] and thus go to very low temperatures [S5]. All details on the KA2D system are provided in the main text and the SM of Ref. [S5].

The isoconfigurational average has been calculated over $N_R = 20$ independent replicas, and $N_S = 1500$ structures are provided for training. Each structure consists of 600 type 1 particles, which thus also yields a total of $N_d \approx 10^6$ data points for training.

For both the KA and the KA2D system, the structural relaxation time τ_α is defined as,

$$F_s(\tau_\alpha) = e^{-1}, \quad (\text{S1})$$

where $F_s(t) = \langle \sum_i \cos(q\Delta r_i(t)) \rangle$ is the incoherent intermediate scattering function with $q \approx 2\pi/\sigma$ and $\Delta r_i(t) = |\mathbf{r}_i(t) - \mathbf{r}_i(0)|$. The structural relaxation time τ_α therefore describes the time scale on which the particles move on average approximately one particle diameter σ .

Equilibrated structures and labels calculated from the relaxation dynamics for both KA and KA2D will be uploaded in combination with this roadmap for future research (see [10.5281/zenodo.10118191](https://doi.org/10.5281/zenodo.10118191)).

C. Propensity (labels)

The input of each ML method are the positions of each particle in a given structure or derived structural descriptors. The labels that are used for training and testing are dynamical quantities characterizing structural relaxation at various time scales t . As introduced in Sec. III of the main text we use propensities to remove any effects of the initial velocities.

For each figure in the main text, except Fig. 5, the results are based on the propensities of displacement, $\mathcal{R}_i = \langle \Delta r_i(t) \rangle_{\text{iso}} = N_R^{-1} \sum_{k=1}^{N_R} |\mathbf{r}_i^k(t) - \mathbf{r}_i(0)|$, where $\mathbf{r}_i^k(t)$ denotes the position of particle i in replica k . In Fig. 5 we have reported results for the bond-breaking propensity of particle i , $\mathcal{C}_B^i(t) = \langle n_t^i / n_0^i \rangle_{\text{iso}}$. Here, n_t^i describes the number of original nearest neighbours particle i still has after a time t and n_0^i is the initial number of neighbours [S6]. The quantity therefore decays from $\mathcal{C}_B^i(t=0) = 1$ to $\mathcal{C}_B^i(t \rightarrow \infty) = 0$. Details and parameters are described in the SM of Ref. [S5].

D. BOTAN

The model architecture of BOTAN [S2] is based on that of the original GNN proposed by Bapst *et al.* [S7]. The difference is that BOTAN is equipped with an additional MLP layer as a decoder for targets on graph edges. BOTAN is implemented as an ‘‘interaction network’’ [S8] consisting of a pair of two-layer MLPs with the size of (64, 64). Each of the MLPs is connected to nodes and edges, respectively, and performs message passing with the other.

The input graphs are constructed by choosing each particle as a node of the graph, and ‘‘connecting’’ nearest-neighbour pairs by edges. The threshold distance for the nearest neighbours are set to $2.0\sigma_{11}$ (both for KA and 2DKA). For recursive message passing, the MLPs repeat the message passing iteratively 7 (KA) or 8 (2DKA)

times, but the results are not strongly affected by slight changes in this number. The target quantity on the edges is, similarly to Ref. [S2], the isoconfigurational average of distance change between particle pairs which are connected via the edges. The loss function is defined as $\mathcal{L} = 0.4\mathcal{L}_{\mathcal{N}} + 0.6\mathcal{L}_{\mathcal{E}}$, where $\mathcal{L}_{\mathcal{N}}$ and $\mathcal{L}_{\mathcal{E}}$ denote the MSE losses for node target (particle propensity) and edge target (propensity for pair-distance change), respectively. Details and visualisation of the method are presented in Ref. [S2].

The BOTAN GNN model contains about 54,200 weight parameters (both for 2D and 3D). The training is started by using pretrained model parameters (optimized to yield the edge target quantities in advance), and is performed over 1,000 epochs on 400 snapshots (KA) and over 2,000 epochs on 1,500 snapshots (2DKA), respectively. The batch size is fixed to one graph. The training has been conducted separately for each temperature and time, by using the Adam optimizer with learning rates of 10^{-4} (KA) and 2×10^{-4} (2DKA). The overall training time amounted to 4 hours (KA) and 11.5 hours (2DKA) on an NVIDIA A100 Tensor Core GPU (40GB SXM) on Aquarius subsystem of Wisteria/BDEC-01 Supercomputer at Information Technology Center, University of Tokyo, with the host CPUs being 2 Intel Xeon Platinum 8360Y. These training times do not include the elapsed time of the data loader. The code and pretrained model parameters for 3D are provided on a GitHub repository (https://github.com/h3-Open-BDEC/pyg_botan).

E. CAGE

As described in Ref. [S9], we train a simple Ridge regression model to fit the dynamics, based on the structure of both the initial state and the cage state. The cage state is defined as the average position of particles before any rearrangement occurs. To obtain the cage state, we perform a Monte Carlo simulation where all the particles are confined to a sphere of radius r_c^α . For both glassy systems we use $r_c^\alpha = 1.25\sigma_{\alpha\alpha}$ with $\sigma_{\alpha\alpha}$ the particle diameter of type α .

To capture the local structure, we use recursive, rotationally invariant parameters that capture both the radial density, as well as the n -fold symmetry in various shells around each particle. For the 3D system, we use exactly the same set of descriptors as in Ref. [S9]. For the 2D system we use the same density parameters as for 3D [S9], however the angular parameters are altered to better reflect 2D symmetry. Specifically, to capture the n -fold symmetry for the 2D system we use

$$\Phi_i^l(r, \delta) = \sqrt{\phi_i^l \cdot (\phi_i^l)^*}$$

with

$$\phi_i^l(r, \delta) = \frac{1}{Z} \sum_{i \neq j} e^{-\frac{(r_{ij}-r)^2}{2\delta^2}} e^{il\theta_{ij}},$$

and

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

where θ_{ij} is the angle between a fixed axis (e.g., x - or y -axis) and the bond joining the i^{th} particle with a particle j . In both 2D and 3D we consider $l \in \{1, 12\}$.

Combining both the radial density and the angular parameters, we obtain a total of 462 parameters for the KA2D system:

- 294 parameters that described the radial density up to the 5th minimum in the pair correlation function (which is located at $4.8\sigma_{11}$).
 - 60 equally spaced in the interval $r/\sigma_{11} \in (0.5, 2.0]$ with $\delta = 0.025$.
 - 20 equally spaced in the interval $r/\sigma_{11} \in (2.0, 3.0]$ with $\delta = 0.050$.
 - 18 equally spaced in the interval $r/\sigma_{11} \in (3.0, 4.8]$ with $\delta = 0.100$.

Note that the radial density functions are type specific, such that in the ternary system the total number of functions above is multiplied by three.

- 168 parameters that capture the n -fold symmetry up to the second minimum of the pair correlation function (located at $2.3\sigma_{11}$).
 - 14 equally spaced in the interval $r/\sigma_{11} \in [1.0, 2.3]$ with $\delta = 0.1$ and $l \in [1, 12]$.

Note that the radial functions do not take the particle species into account.

As described in Ref [S9], for the 3D system we obtain a total of 366 parameters. In addition to the zeroth order generation parameters, for both the 2D and 3D system, we additionally include two generations of structural parameters that are iteratively averaged over the nearest neighbours (see Ref. [S10]).

The dataset on which we train our Ridge regression thus includes three generations of structural descriptors for both the initial structure, as well as the cage state. Additionally, for each particle we include the distance between the cage state and the initial positions $\Delta r_i^{\text{cage}} = |\mathbf{r}_i^{\text{cage}} - \mathbf{r}_i^{\text{init}}|$. This means that the local structure of each particle in the 2D system is described by a total of 2197 parameters, and in the 3D system a total of 2773 parameters.

To predict the dynamics, we standardise the data and then train the Ridge regression model on 300 (KA2D) and 100 (3D KA) snapshots, respectively. The only free parameter that is tuned for this ML method is the regularization parameter α , which sets the strength of the penalty for large weights in the Ridge regression. We

train the model on various values of $\alpha \in \{10^{-5}, 10^5\}$, and then choose the α that yields the highest correlation. The training time per timestep in both 2D and 3D is less than 10 minutes on a standard CPU.

F. GlassMLP

The methodology employed in this manuscript corresponds exactly to the procedure described in detail in Ref. [S5], including all training meta data. In particular, we calculate a set of M_S physics-inspired descriptors, including the coarse-grained local density, $\bar{\rho}_{L,\beta}^i = \sum_{j \in N_\beta^i} e^{-r_{ij}/L}$, which is coarse-grained over a distance L by summing over all N_β^i particles of type β within distance $r_{ij} = |\mathbf{r}_i^{\text{inh}} - \mathbf{r}_j^{\text{inh}}| < 20\sigma_{11}$ of particle i . Importantly, particle positions are extracted from the inherent structures $\mathbf{r}_i^{\text{inh}}$, and not directly from the thermal structure. Similarly, we include the potential energy, perimeter of the Voronoi cell and local variance of potential energy (see Ref. [S5] for details).

Different from Ref. [S5] we additionally include one further structural descriptor, based on the distance between the inherent and the thermal states of particle i , $\Delta r_i^{\text{inh}} = |\mathbf{r}_i^{\text{inh}} - \mathbf{r}_i^{\text{th}}|$: $\bar{\Delta r}_{L,\beta}^i = \sum_{j \in N_\beta^i} \Delta r_j^{\text{inh}} e^{-r_{ij}/L}$. This descriptor does not affect the long-time prediction and was therefore discarded in Ref. [S5], but is reintroduced here mainly to improve the short-time performance of GlassMLP.

To predict propensity in the KA (3D) system, GlassMLP has in total 618 free parameters. In the ternary KA2D system, the number is slightly increased (765). The training just takes 2-5 minutes with an octa-core CPU or a Laptop GPU (NVIDIA T600 Laptop).

G. SE(3)

The model is exactly that presented as the main model of Ref. [S11] (version 2, august 2023).

We recall the main features for completeness and to avoid possible confusion. In the SE(3) GNN we use 8 layers. The structure of each layer is the same: 8 channels for each l component, with $l = 0, 1, 2, 3$ (namely, $l_{\text{max}} = 3$). The input data consists of the thermal positions (not quenched to the IS) and of the particles potential energy (but computed from the inherent state's particles positions). We predict simultaneously propensities for all particle types and time steps. The connectivity graph is computed using a threshold distance of $2.0\sigma_{11}$. The input node features are the particle type (one-hot encoded) concatenated with the potential energy (of the IS). For the 2D case, we introduce an artificial third component to the position of all particles ($\forall i, z_i = 0$) and used spherical harmonics embedding. This is overkill but is a quick way to adapt the 3D scheme to 2D data.

The radial-encoding MLP encodes the radius on (Bessel) basis functions using 10 basis vectors, followed by a hidden layer of 16 neurons, and uses a dropout of 0.3. We use batch normalization between each convolution layer. We minimise the MSE loss with a L^2 regularization coefficient $\beta = 10^{-7}$, using a batch size of 2 graphs (8 in 2D), a learning rate of 10^{-3} with Adam ($\beta_1 = 0.99$). As opposed to Ref. [S11], we do not use a validation set in this manuscript and therefore do not perform early stopping. We simply take the last epoch model (the accuracy and losses are basically flat when we interrupt learning).

The model has 52,660 parameters for 3D data (respectively 53,394 in 2D). We perform 200 epochs (100 in 2D), which takes approximately 11 hours (27 hours in 2D) using an NVIDIA Tesla V100 (32GB) (the CPUs on the node are 2 Intel Xeon Gold 6148 20 cores (40 threads) at 2.4 GHz (Skylake)). Note that we train simultaneously the 10 time steps (6 for KA2D) of a given state point (temperature), therefore the training does not have to be repeated for each timescale.

H. SBO

Unsupervised learning of local structure fluctuations is carried out for the 3D KA samples along the lines of Ref. [S12], using the `partycls` Python package [S13]. Namely, we characterise the local structure around particle i using the smooth bond-order (SBO) descriptor

$$X^{\text{SBO}}(i) = (Q_0^S(i), \dots, Q_{l_{\text{max}}}^S(i)), \quad (\text{S2})$$

where Q_l^S are smoothed bond-orientational invariants of order l computed over the first coordination shell of particle i , see Refs. [S12] for full details. The maximum order l_{max} is equal to 8. A PCA is then carried out for each chemical species, to identify the directions in the descriptor space that capture the largest structural fluctuations. In particular, the projection on the first principal component, $\tilde{X}_1^{\text{SBO}}(i)$, is a measure of structural heterogeneity that is correlated to some extent with the dynamics of glassy binary mixtures [S12]. The correlation between the first projection of the SBO descriptor and the propensity of motion is calculated directly on the test dataset, since there is usually no separate training stage in unsupervised learning. Finally, along with the bare SBO descriptor, we consider a coarse-grained (CG) version

$$\tilde{X}_i^{\text{CG}} = \frac{\sum_j \tilde{X}_j \cdot w(r_{ij}; L)}{\sum_{j=1}^N w(r_{ij}; L)},$$

where the sum over j runs over the particles of the same species as i . As in Ref. [S14] we use an exponential function $w(r; L) = e^{-r/L}$, and we set $L = 1$.

I. Other structural descriptors

In addition to these models we have also included several traditional structural descriptors.

Density:

The density is connected to the structural input of GlassMLP (see definition Sec. [IF](#) above). In particular, we have used $\bar{\rho}_{L=5,\beta=\text{all}}^i$, i.e. coarse-grained over a distance $L = 5$ considering particles of all types. The only difference to GlassMLP is that we have evaluated the density in the thermal states.

Potential energy $\bar{E}_{\text{pot}}^{\text{inh}}$:

The potential energy in the inherent state, $\bar{E}_{\text{pot}}^{\text{inh}} = \bar{E}_{L=5,\beta=\text{all}}^i = \sum_{j \in N_\beta^i} E^j e^{-r_{ij}/L} / \bar{\rho}_{L,\beta}^i$, is similarly part of the structural descriptors which is used as input for GlassMLP [[S5](#)]. Here, $E^i = \sum_{j \neq i} V(r_{ij})/2$ is the potential energy of particle i .

Bond-order 3D (Ψ_6):

The bond-order descriptor Ψ_6^i of particle i for the KA system in three dimensions is defined via the complex coefficient [[S15](#)]

$$q_{l,m}^i = \frac{1}{N_b} \sum_{j \in N_b} Y_l^m(\mathbf{r}_{ij}), \quad (\text{S3})$$

where N_b is the number of bonds of particle i , defined as all neighbours within a cutoff of $r_{\text{cut}} = 2.0\sigma_{11}$, and $Y_l^m(\mathbf{r}_{ij})$ are the spherical harmonics. Using q^i we define,

$$\Psi_l^i = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{m=l} |q_{l,m}^i|^2}. \quad (\text{S4})$$

As before, we also further coarse-grain the descriptor, $\bar{\Psi}_{6,L=5,\beta=\text{all}}^i = \sum_{j \in N_\beta^i} \Psi_6^j e^{-r_{ij}/L} / \bar{\rho}_{L,\beta}^i$.

Bond-order 2D (Ψ_4):

For the KA2D system, we slightly adapt the definition,

$$q_l^i = \frac{1}{N_b} \sum_{j \in N_b} e^{il\theta_{ij}}, \quad (\text{S5})$$

which is strongly related to the n -fold symmetry defined in Sec. [IE](#). From this we calculate,

$$\Psi_l^i = \sqrt{q_l^i \cdot (q_l^i)^*}, \quad (\text{S6})$$

which is subsequently coarse-grained in the same way as defined in the 3D case.

Tanaka's Θ order parameter:

In Ref. [[S16](#)] Tong and Tanaka proposed the usage of a structural descriptor Θ^i , which quantifies the strength of local packing around particle i (see Ref. [[S16](#)] for details

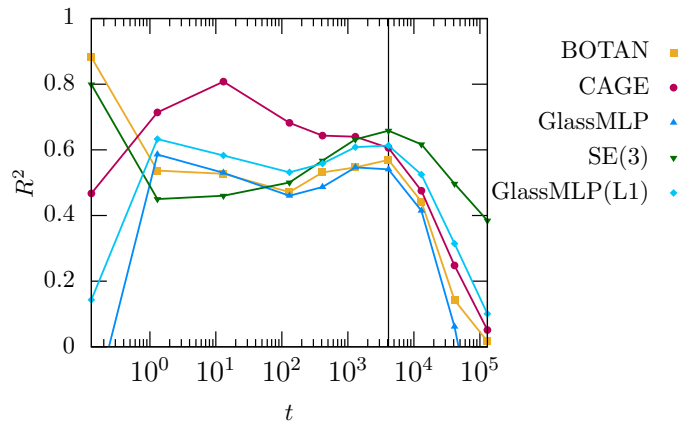


FIG. S1. Coefficient of determination $R^2 = 1 - \sum_i (\mathcal{R}_i^{\text{MD}} - \mathcal{X}_i^{\text{ML}})^2 / \text{var}(\mathcal{R}_i^{\text{MD}})$ between the model predictions and the ground truth in the KA system (3D) at $T = 0.44$ for various ML models. The dynamical variable is the propensity of displacements. The vertical line marks the structural relaxation time τ_α . GlassMLP(L1) denotes the GlassMLP model [[S5](#)] trained without the additional terms in the loss function using only the L^1 norm.

and definitions). We similarly coarse-grain this descriptor $\bar{\Theta}_{L=5,\beta=\text{all}}^i = \sum_{j \in N_\beta^i} \Theta^j e^{-r_{ij}/L} / \bar{\rho}_{L,\beta}^i$.

In a 2D system of polydisperse harmonic spheres, Ref. [[S16](#)] reports a Spearman's rank correlation coefficient of around $\rho_S \approx 0.9$ for the coarse-grained $\bar{\Theta}$ with propensity. ρ_S is usually strongly related to Pearson correlation ρ_P . In this roadmap, we show that the performance is significantly reduced when applied to systems which are not prone to crystallization. Additionally, some of the performance difference might emerge from calculating Pearson correlation over all particles independent of their radii in Ref. [[S16](#)], as opposed to making independent predictions for each particle type as done in this manuscript. This emphasises the importance of using standardized datasets and performance metrics to quantify and validate the performance of newly developed descriptors and techniques.

II. ADDITIONAL BENCHMARKING

In the following, we will discuss additional metrics and ways to visualise the performance of the various methodologies which have been benchmarked intensively in Sec. V of the main text.

A. Coefficient of determination R^2

The coefficient of determination $R^2 = 1 - \sum_i (\mathcal{R}_i^{\text{MD}} - \mathcal{X}_i^{\text{ML}})^2 / \text{var}(\mathcal{R}_i^{\text{MD}})$ is a popular measure to quantify the quality of a fit, including, in particular, supervised ML models. In Fig. [S1](#) the coefficient of de-

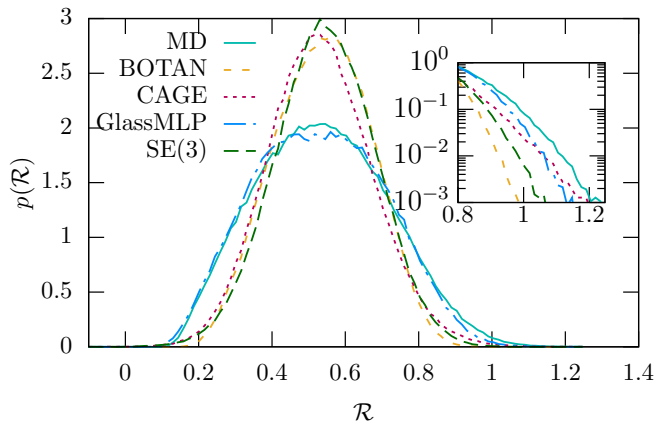


FIG. S2. Probability distribution of predicted and simulated propensities at $t = \tau_\alpha$ for the KA system (3D). The dynamical variable is the propensity of displacements \mathcal{R} . The inset shows a zoom to the tail, with logarithmic y-axis.

termination is shown for exactly the same data as was analysed with the Pearson correlation in Fig. 3. It can be observed that, in general, the two measures ρ_P and R^2 do not show very different results. In particular, SE(3) still features the best performance at $t = \tau_\alpha$. The most pronounced difference is that GlassMLP performs worse when measured with R^2 instead of ρ_P . This can be explained by the usage of a loss function which contains additional terms on top of the standard L^1 or L^2 norm [S5]. To validate this explanation we show the results for GlassMLP(L1) trained without these additional contributions and find results significantly better than for the original GlassMLP (see Fig. S1).

B. Probability distribution of propensities

Another interesting observable is the probability distribution of the propensities of displacement, $p(\mathcal{R})$, which can be extracted both from the MD simulations and from the ML models. A similar analysis has been performed in Ref. [S5]. As can be seen in Fig. S2 the propensities predicted by the models tend to underestimate the variance of the distribution. This observation can be rationalized, since predicting outliers could be very costly in the loss function during training, hence networks tend to predict small variances. The only network which agrees quite well to the MD prediction is GlassMLP, which has been constructed to adapt the variance by an explicit contribution in its loss function [S5]. However, as can be seen in the inset of Fig. S2, even for GlassMLP the tail of strongly moving particles is not perfectly reproduced. Focusing on this tail would likely significantly worsen the overall performance of the model.

C. Scatter plots of propensity predictions

The Pearson correlation ρ_P or the coefficient of determination R^2 are very good measures to quantify the performance of a network using a scalar quantity (see Figs. 3-6 in the main text and Fig. S1). However, they obviously cannot capture all details of the connection between the true and predicted propensities. In Fig. S3 we therefore show a 2D histogram visualizing this connection in greater detail. The overall impression of these figures is very similar to what we have concluded before. The most obvious difference between the methods is that GlassMLP is symmetric around the straight line, while the other techniques are slightly asymmetric. This observation is connected to the differences in the probability distribution as discussed in the previous paragraph.

D. Snapshots

The best way to visualise the predictions are snapshots of the amorphous structure, where each particle is coloured according to its propensity \mathcal{R} . In Fig. S4 we show an exemplary structure and the predicted propensities of displacement for various different models. Generally, strong correlations between the models and the MD simulations can be observed, as was expected from the high Pearson correlations. The strongest visible difference between MD and the ML models is that the latter are much smoother, in particular the ones based on GNNs (BOTAN and SE(3)).

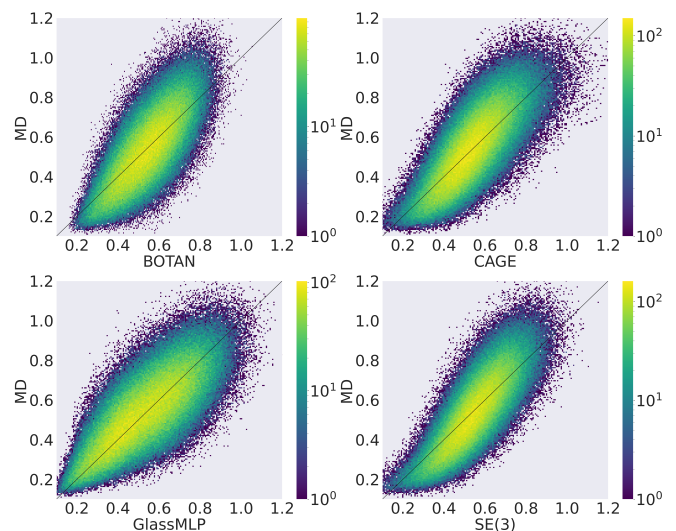


FIG. S3. 2D histograms of particle propensity \mathcal{R}_i to compare different ML techniques to the MD ground truth in the KA system (3D, $t = \tau_\alpha$). Each entry in the histogram represents one particle, i , where the x/y-coordinates correspond to the propensities as predicted by ML and from MD simulations, respectively.

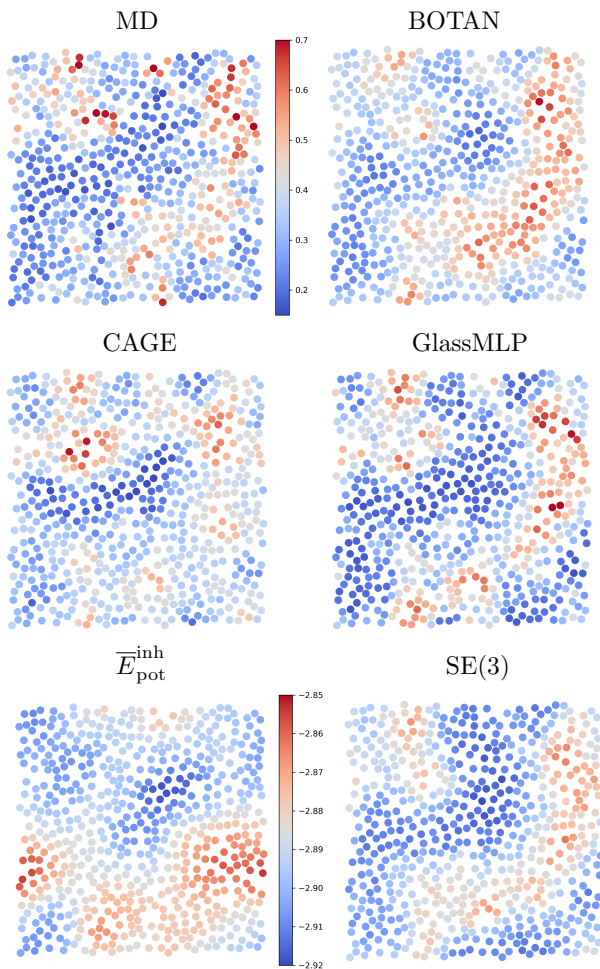


FIG. S4. Snapshots of an exemplary configuration of the KA2D system (shown are only type A particles). Each particle is coloured according to its propensity of displacement \mathcal{R} at $t = \tau_\alpha$, as calculated from MD simulations or predicted from various ML techniques (see Fig. 3 in the main text for the respective Pearson correlations). Each snapshot has the same color bar, except $\bar{E}_{\text{pot}}^{\text{inh}}$.

E. Cross correlations

To study connections between the different models, we also investigate the cross correlations between the different ML techniques and other structural descriptors. Unsurprisingly, the strongest cross correlations are between the various ML models, since they all correlate strongly with propensity. The strongest correlation can, in fact, be detected between SE(3) and GlassMLP (see Fig. S5). This can be rationalized by the observation that both techniques also have the strongest correlation with the coarse-grained potential energy, $\bar{E}_{\text{pot}}^{\text{inh}}$. This could be expected since both techniques receive the potential energy as explicit input.

Another strong cross correlation is observable between

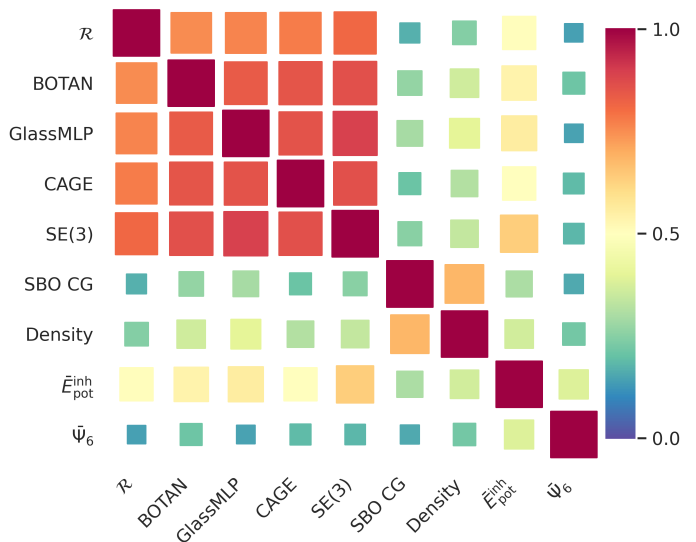


FIG. S5. Pearson cross-correlations $\rho_P(X, Y)$ between different ML models and structural descriptors in the KA system (3D) at $t = \tau_\alpha$.

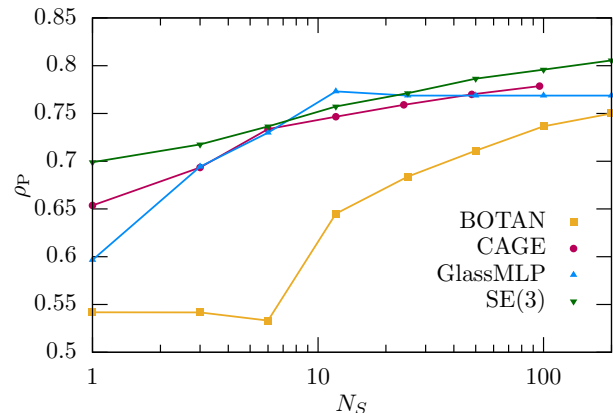


FIG. S6. Performance of network, as quantified by the Pearson correlation coefficient ρ_P , trained with differently sized training sets in the KA system (3D) at $t = \tau_\alpha$. N_S denotes the number of samples, where each sample consists of $N = 3277$ particles.

the unsupervised SBO CG technique and density, showing that these two descriptors are strongly linked. This indicates that the strongest asymmetries in the structure emerge from density fluctuations.

F. Learning curves

An important feature of ML models is also how they cope with limited training data. In standard fitting procedures, one would usually expect that the more fitting parameters a model has, the more training data is required to achieve a certain accuracy. This 'rule of thumb'

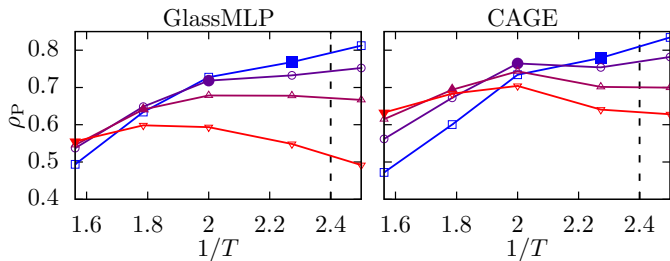


FIG. S7. Transferability in temperature T of trained networks in the KA system (3D) at the structural relaxation time τ_α . The figures is identical to Fig. 6 in the main text, just featuring an additional temperature $T = 0.4$.

has, however, been repeatedly questioned for deep networks, which often perform well even when overparameterized. When investigating Fig. S6 we observe that indeed SE(3) already performs excellent, even with only $N = 3277$ data points, despite having $> 50k$ fitting parameters. Furthermore, the increase in performance is very continuous, implying that the network might perform even better with > 400 independent structures (which was the maximum of the dataset available). The learning curves of CAGE and BOTAN closely resembles SE(3) just with a very small reduction in performance.

Contrarily, the learning curve of GlassMLP starts at a significantly smaller Pearson correlation, but quickly catches up with SE(3) to slightly surpass its performance at around $N_S = 10$ (i.e. 32770 data points). However, the performance of GlassMLP does not further improve for larger $N_S > 20$. It might be necessary to increase the number of structural input descriptors to improve GlassMLP in situations where much training data is available. It should be mentioned that there is an important difference between the training procedures of SE(3) and the other models. While GlassMLP uses for $N_S = 1$ only twice the number of epochs for training (i.e. $N_e = 600$), since no improvement in performance can be observed beyond this point, SE(3) requires a significant increase in epochs for small N_S : $N_e(N_S) = 8 \cdot 10^4 / N_S$. This implies that for $N_S = 1$, the training time of GlassMLP is roughly 5 seconds, while for SE(3) it remains roughly 11 hours.

G. Transferability

We have investigated in detail the transferability of trained networks in the benchmarking section V of the main text. Here, we study for two models the performance of GlassMLP and CAGE when applied to an even lower temperature $T = 0.4$, taken from the dataset presented in Ref. [S17]. Different from the main KA dataset we only have $N_S = 12$ individual structures available, but use $N_R = 100$ different isoconfigurational replicas. Due to the larger number of replicas we expect the Pearson correlations to be slightly larger than for $N_R = 32$.

The results in Fig. S7 indeed confirm this expectation, nevertheless, they also highlight the strong transferability of the models. A network trained at $T = 0.44$, with a relaxation time an order of magnitude smaller than at $T = 0.4$, CAGE can predict structural relaxation with a Pearson correlation of up to $\rho_P = 0.84$. We believe that this is an important result towards using transferability to study glassy liquids at extremely low temperature.

H. Bond-breaking propensity in the KA system (3D)

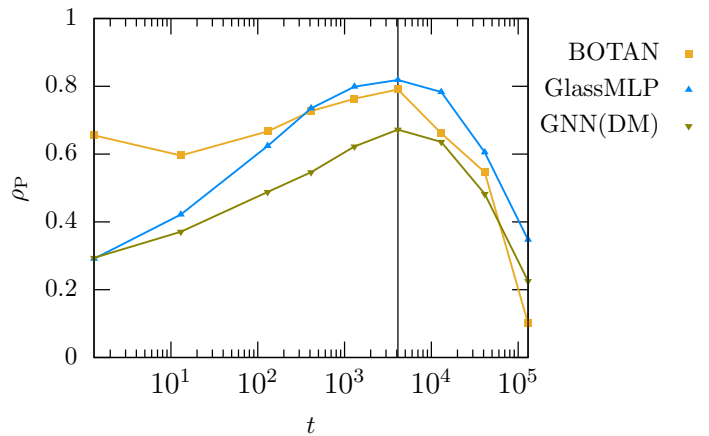


FIG. S8. Pearson correlation ρ_P with the ground truth in the KA system (3D) at $T = 0.44$ for various different techniques. Different from Fig. 3, here the dynamical variable is the bond-breaking propensity. The vertical line marks the structural relaxation time τ_α . GNN(DM) refers to the original GNN [S7].

Finally, we also briefly investigate the performance of ML models and other structural descriptors in predicting the bond-breaking propensity in the KA system (see Fig. S8). While the predictability for small times is much weaker than for the propensity of displacements, \mathcal{R} , analysed in Fig. 3, we observe better performance for longer times around the structural relaxation time scale $t = \tau_\alpha$. This observation is identical to the discussion in Sec. V of the main text after Fig. 5 for the bond-breaking propensity in the KA2D system. We further find that both BOTAN and GlassMLP significantly outperform the original GNN proposed by DeepMind ([S7], GNN(DM)). This shows the great advancements in the field within the past three years. We hope that this roadmap further fuels the development and application of ML techniques to analyse glassy liquids.

SUPPLEMENTARY REFERENCES

- [S1] 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).
- [S2] H. Shiba, M. Hanai, T. Suzumura, and T. Shimokawabe, BOTAN: BOnd TArgeting Network for prediction of slow glassy dynamics by machine learning relative motion, *The Journal of Chemical Physics* **158**, 084503 (2023).
- [S3] A. Ninarello, L. Berthier, and D. Coslovich, Models and algorithms for the next generation of glass transition studies, *Phys. Rev. X* **7**, 021039 (2017).
- [S4] L. Berthier, E. Flenner, C. J. Fullerton, C. Scalliet, and M. Singh, Efficient swap algorithms for molecular dynamics simulations of equilibrium supercooled liquids, *Journal of Statistical Mechanics: Theory and Experiment* **2019**, 064004 (2019).
- [S5] 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).
- [S6] B. Guiselin, C. Scalliet, and L. Berthier, Microscopic origin of excess wings in relaxation spectra of supercooled liquids, *Nature Physics* **18**, 468 (2022).
- [S7] 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, *Nature Physics* **16**, 448 (2020).
- [S8] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, C. Gulcehre, F. Song, A. Ballard, J. Gilmer, G. Dahl, A. Vaswani, K. Allen, C. Nash, V. Langston, C. Dyer, N. Heess, D. Wierstra, P. Kohli, M. Botvinick, O. Vinyals, Y. Li, and R. Pascanu, Relational inductive biases, deep learning, and graph networks (2018), [arXiv:1806.01261](https://arxiv.org/abs/1806.01261).
- [S9] R. M. Alkemade, F. Smalenburg, and L. Filion, Improving the prediction of glassy dynamics by pinpointing the local cage, *The Journal of Chemical Physics* **158**, 134512 (2023).
- [S10] E. Boattini, F. Smalenburg, and L. Filion, Averaging local structure to predict the dynamic propensity in supercooled liquids, *Phys. Rev. Lett.* **127**, 088007 (2021).
- [S11] F. S. Pezzicoli, G. Charpiat, and F. P. Landes, SE(3)-equivariant graph neural networks for learning glassy liquids representations, *arXiv preprint arXiv:2211.03226* (2022).
- [S12] D. Coslovich, R. L. Jack, and J. Paret, Dimensionality reduction of local structure in glassy binary mixtures, *J. Chem. Phys.* **157**, 204503 (2022).
- [S13] J. Paret and D. Coslovich, partycls: A python package for structural clustering, *JOSS* **6**, 3723 (2021).
- [S14] 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).
- [S15] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Bond-orientational order in liquids and glasses, *Phys. Rev. B* **28**, 784 (1983).
- [S16] H. Tong and H. Tanaka, Structural order as a genuine control parameter of dynamics in simple glass formers, *Nature communications* **10**, 5596 (2019).
- [S17] 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, *The European Physical Journal E* **41**, 1 (2018).