Machine learning caging order parameters in glasses

Kaihua Zhang,1,Xinyang Li,2,3,Yuliang Jin,2,3,† and Ying Jiang1,4,‡

1School of Chemistry, Beihang University, Beijing 100191, China
2CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China
3School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
4Center of Soft Matter Physics and Its Applications, Beihang University, Beijing 100191, China

Non-equilibrium phase transitions in glassy systems are often indicated by a dramatic change of dynamics, accompanied by subtle and ambiguous structural signatures. This fact has motivated a number of recent studies attempting to pinpoint predictors that correlate structural order to dynamics in glasses, via the application of modern machine learning techniques. Here we develop a general machine learning approach, based on a two-level nested neural network, which autonomously extracts glass order parameters characterizing caging dynamics. Combining machine learning with finite-size scaling analyses, we can identify, and distinguish between, first- and second-order non-equilibrium phase transitions, demonstrated by studying melting and Gardner transitions in a simulated hard sphere glass model. Our machine learning results also suggest that the liquid-to-glass transition, as widely accepted, is a smooth crossover rather than a sharp phase transition. Our approach makes use of the power of neural networks in learning hidden dynamical features of different phases, bypassing the difficulties in defining structural order in disordered systems.

Introduction

The essential idea of Landau’s celebrated theory of phase transitions is that the two phases around a transition can be distinguished by suitable order parameters. For example, a liquid-to-crystal transition is characterized by order parameters that quantify the corresponding crystal structure. However, the problem is not as straightforward as that for non-equilibrium transitions related to glasses (aka amorphous solids), such as glass (liquid-to-glass) transitions, melting (glass-to-liquid) transitions, and glass-to-glass transitions. Around these transitions, dynamics change significantly—e.g., the relaxation time can increase by more than ten orders of magnitude over a small range of temperatures around the glass transition [11]. Meanwhile, the two phases on both sides of the transition appear to be “disordered”. Great efforts have been devoted to finding a proper glass order parameter [2,3], based on static structures or energetic information. Candidates include the bond orientation order [4], the potential energy [5,6], soft modes [7], the tetrahedral order [8,9], the icosahedral order [10], the local five-fold symmetry [11–13], the local packing capability [14–15], and the medium-range order [16]. Unfortunately, a widely accepted static order parameter, universal for different glassy systems, is still unavailable.

While the descriptors mentioned above are motivated by insights in physics, in recent years, researchers have attempted to address the conundrum by taking advantage of the power of machine learning techniques in extracting useful information from complex, massive data. A parameter called “softness” was learned to predict structural flow defects and plasticity in amorphous solids [18–21]. Graph neural networks (GNN) were used to predict long-time dynamics (up to the diffusive regime) from the graph structure of initial particle positions [22]. Inspired by the GNN method, new structural descriptors were designed by recursively incorporating averaged features from neighbour shells [23]. Unsupervised machine learning methods were also developed to identify different clusters of particles based on their static correlations [24–25].

In parallel to the static approach of structural order parameters, alternatively, one may consider glass caging order parameters defined purely based on dynamics (see below for a detailed description) [26–27]. Compared to structural order parameters, the caging order parameters are often more generic. For example, spin and structural glasses share very similar dynamics, but their structures are obviously different. In fact, in spin glasses, it is standard to define and analyze overlapping order parameters [28], which play a similar role as the caging order parameters in structural glasses. The caging/overlapping order parameters are at the core of mean-field theories of glass transitions, such as the mode-coupling theory [29–31], the replica theory [26–28], and the random first-order phase transition theory [32–34].

Here, by using machine learning methods, we study in detail three transitions in a hard sphere (HS) glass model [34]: the melting transition [35–36], the Gardner transition (a glass-to-glass transition) [37–38], and the glass transition. The outcome is three-fold. (i) Inspired by mean-field glass theories [26–33], we design a two-level nested neural network (NNN), which takes as input replicated configurations generated by molecular simulations. We show that, once well trained, the networks at the first level can provide caging order parameters that are suitable for characterizing the specific transition, and the machine-defined order parameters are consistent with physically defined parameters. This feature distinguishes our dynamics-oriented machine learning method...
Any state of the system is described by two thermodynamic parameters, the volume fraction $\varphi$ and the reduced temperature $\hat{T}$. The Carnahan-Stirling (CS) equation of state (EOS) [44] well captures the relationship between $\varphi$ and $\hat{T}$ of liquid states [17]. The mode-coupling theory (MCT) transition point, $T_{\text{MCT}} \approx 0.044$ (or $\varphi_{\text{MCT}} \approx 0.594$), was estimated in [17], below which equilibrium becomes difficult in ordinary molecular dynamics (MD) simulations. Following convention, the intersection between liquid and glass EOSs is defined as the glass transition point $\{\varphi_g, \hat{T}_g\}$. The glass transition point is protocol-dependent and, therefore, not unique. Two kinds of glasses, ordinary and ultra-stable glasses, are generated by our molecular simulations.

(i) To study the glass transition, we consider ordinary glasses. Ordinary glasses are created by MD compression (see Methods) from low-density liquid configurations, with a fixed compression rate $\Gamma$. In this study, we focus on the case of $\Gamma = 10^{-3}$, corresponding to a glass transition temperature $\hat{T}_g \approx 0.047$ (or glass transition density $\varphi_g \approx 0.58$) close to $T_{\text{MCT}}$.

(ii) To study melting and Gardner transitions, we consider ultra-stable glasses. Deeply supercooled liquid states are prepared by using an efficient swap Monte Carlo (MC) algorithm [34] (see Methods); once the initial states are obtained, we switch to regular MD (without swap) to simulate follow-up dynamics. These deeply supercooled liquid states have extraordinarily large structural relaxation ($\alpha$-relaxation) time $\tau_\alpha$ in the MD time unit, much larger than our MD simulation time window. Under MD compression ($\Gamma > 0$) or decompression ($\Gamma < 0$), where $1/|\Gamma| \ll \tau_\alpha$, the system falls out of equilibrium and becomes a ultra-stable glass [45].

Two instabilities – the melting and the Gardner transitions – occur if one decompresses or compresses an ultra-stable HS glass. Under decompression, the glass is effectively “heated” up and eventually melts into a liquid at a melting temperature $T_m(\hat{T}_g) > \hat{T}_g$. Although glass melting is a non-equilibrium procedure by definition, previous experiments [46] and simulations [35] showed that, this procedure in ultra-stable glasses is very similar to the melting of crystals, which is a standard first-order phase transition. As can be seen in Fig. 1, the EOS of an ultra-stable glass ($\hat{T}_g \ll T_{\text{MCT}}$) displays enormous overshooting over the liquid EOS and the two EOSs are connected by an abrupt jump. In contrast, the EOS of a poorly annealed ordinary glass ($\hat{T}_g \approx T_{\text{MCT}}$) merges smoothly to the liquid one. Recent studies [36, 37] suggest that the discontinuous melting of ultra-stable glasses is a vestige of a hidden first-order phase transition, which takes place in two coupled replicas of the system.

On the other hand, a Gardner transition [38, 39] is expected to occur at $\hat{T}_g(\hat{T}_G) < \hat{T}_g$ if an ultra-stable HS glass is compressed. The Gardner transition separates the stable glass (at $\hat{T} > \hat{T}_G$) and the marginally stable glass (at $\hat{T} < \hat{T}_G$) phases. It is predicted to be a second-order phase transition in large dimensions by the

![Phase diagram of a poly-disperse hard sphere glass](image)

**FIG. 1.** Phase diagram of a poly-disperse hard sphere glass [17]. The green, blue and red lines represent the CS liquid EOS, the melting line that separates liquid and glass phases, and the Gardner line that separates stable and marginally stable glass phases. The orange star represents the MCT transition point. The system is initially equilibrated at $\hat{T}_g$ (green squares), and then evolves following the glass EOSs (dotted black lines) under compression or decompression. In this study, we focus on the ultra-stable case of $\hat{T}_g = 0.033$ (solid black line). The melting points (blue diamonds) are determined at the maximum of pressure fluctuations (see Ref. [17] for the determinations of other points and lines). Typical particle trajectories are plotted to show the diffusive motion in liquids, a confined cage in stable glasses, and the split of the cage into sub-cages in marginal glasses (the three sub-cages are visualized by trajectories of the same particle in three replicas, which are compressed from the same initial configuration at $\hat{T}_g$).

from previous ones based on pure structures [18–25]. (ii) The network at the second level predicts statistically to which phase the sample belongs. Applying finite-size scaling analyses to the predictions [39–42], we determine that the melting transition is first-order and the Gardner transition is second-order. The estimated correlation length critical exponent $\nu$ of the Gardner transition agrees with previous results [42, 43]. (iii) Unlike in the cases of melting and Gardner transitions, the machine estimated glass transition point depends systematically on algorithm parameters, suggesting that the transition from liquids to glasses is a smooth crossover rather than a well-defined phase transition.

**Results**

**Glass, melting and Gardner transitions in a hard sphere glass**

We study a polydisperse HS glass model (see Methods), whose phase diagram is presented in Fig. 1 [17].
FIG. 2. **Schematic diagram of the machine learning method.** The core of the method is a two-level NNN. The input data to the $i$th small network on the first-level are $\{\Delta_{i1}, \Delta_{i2}, \ldots, \Delta_{iN_r-1,N_r}\}$ (or $\{u_{i1}, u_{i2}, \ldots, u_{iN_r-1,N_r}\}$ for the Gardner transition, see Methods), which are computed from $N_r$ replicated configurations generated by MD simulations. The first-level networks can autonomously extract the caging order parameter $q_i$ from the input data. The output of the big FNN on the second-level is a classification of phases.

mean-field glass theory [27, 38]. Evidence of the Gardner transition in physical dimensions (2D and 3D) has been reported in a number of simulations [17, 47, 48] and experiments [49, 51]. A fixed point is found by field-theory calculations, suggesting that the transition survives in low dimensions [43]. The existence of a Gardner transition in 3D ultra-stable HS glasses is supported by a recent numerical study, which combines finite-time-finite-size analyses with machine learning [42].

In this study, we focus on the melting and Gardner transitions in stable glasses of a glass transition temperature $T_g \approx 0.033$ (or $\varphi_g \approx 0.63$), but we expect our results and conclusions to be general as long as the glass is ultra-stable ($T_g \ll T_{\text{MCT}}$). As demonstrated below, the melting and Gardner transitions in ultra-stable glasses are similar to standard equilibrium first- and second-order phase transitions in the Ising model, with respect to the finite-size scaling behavior (apart from the effect of disorder). We do not study the melting and Gardner transitions in ordinary glasses: The melting of an ordinary glass is nearly reversible to the glass transition, and thus for our purpose, it is sufficient only to consider the latter. The Gardner transition in ordinary glasses is blurred by activated dynamics [17]; as a result, we do not expect to observe critical scalings.

**Caging order parameters**

In the framework of replica glass theory [27], the three transitions are all characterized by caging order parameters. Particles diffuse in the liquid phase, but in the glass phase, they are confined in cages formed by nearby particles (Fig. 1). Thus the average cage size $\Delta$ (see Methods for the definition) can be used as an order parameter to identify glass and melting transitions: $\Delta$ is small and finite in glasses (if activations are neglected as in mean-field theories), while in liquids, $\Delta \to \infty$ in the thermodynamic limit.

The parameter for the Gardner transition is much more complicated. In the marginally stable glass phase, particle cages split into multiple hierarchical sub-cages. According to the mean-field theory [27, 52], the two phases around a Gardner transition are distinguished by, not a single scalar, but a probability distribution function $P(\Delta_i)$ of single-particle cage size $\Delta_i$ (see Methods for the definition). However, in 3D systems, it is more convenient to look at the distribution $P(\chi_i)$ of single-particle caging susceptibility $\chi_i$ (see Methods for the definition), which shows single and double peaks in stable and marginal glasses respectively [42].

**Setup of supervised machine learning and finite-size analyses**

The design of our input data is inspired by the so-called “replica construction”, which is the core of the replica mean-field glass theory [26, 27, 52]. We make $N_r$ identical replicas of an equilibrium sample at $\varphi_g$, and measure their uncorrelated dynamics after compression or decompression (see Methods for details). In the glass phase, the replicated particles are confined in the same glass cage and thus can be grouped into a “molecule”. The caging property of the $i$th particle is characterized by quantities such as the single-particle cage size $\Delta_i$ and the single-particle cage susceptibility $\chi_i$.

According to the above construction, if the initial system has $N$ “atoms”, then the replicated system consists of $N_r$ “molecules”, where each molecule is formed by $N_r$ replicas of the same atom (see Fig. 2). To match such a structure, we design a two-level NNN as demonstrated
FIG. 3. Learning the caging order parameter of the melting transition. (a) Liquid and glass $\hat{T}(\phi)$ EOSs are separated by the melting temperature at $\hat{T}_m \approx 0.064$. The data are obtained by MD simulations of $N = 125$ systems. The caging order parameter $\Delta$ significantly increases when a glass melts into a liquid, which is accurately captured by the linearly rescaled output $\tilde{q}$ from first-level networks, $\Delta = \tilde{q} \equiv c_1 q - c_2$ (see SI Sec. S2 for how $c_1$ and $c_2$ are determined). The two distributions $P(\Delta_i)$ and $P(\tilde{q}_i)$ also coincide in both (b) liquid ($\hat{T} = 0.13$) and (c) glass ($\hat{T} = 0.03$) phases, where $\tilde{q}_i \equiv c_1 q_i - c_2$. The distributions are averaged over $N_i = 2576$ samples. Insets of (b) and (c): correlations between $\Delta_i$ and $q_i$. The estimated Pearson correlation coefficients $r = 1.0$ in both cases. The color bar represents the normalized density of data points.

FIG. 4. Finite-size analysis of the melting transition. (a) Density susceptibility $\chi_\phi$ as a function of $\hat{T}$, for a few different system sizes $N$. (b) Collapse of the $\chi_\phi$ data according to Eq. 1. Best collapsing is obtained by setting $a = 0.8$, $b = 0.5$ and $\hat{T}_c = \hat{T}_m = 0.064$. (c) Machine learning results $P(\hat{T}, N)$ and $1 - P(\hat{T}, N)$ as functions of $\hat{T}$ and (d) $(\hat{T} - \hat{T}_m) N^b$, where $b = 0.496(5)$. The data points are fitted to an error function (lines), $P(\hat{T}, N) = \frac{1}{2} + \frac{1}{2} \text{erf} \left( \frac{\hat{T} - \hat{T}_m(N)}{w(N)} \right)$, where the transition temperature $\hat{T}_m(N)$ (see c-inset) and the width $w(N)$ (see d-inset) are two fitting parameters. The exponent $b = 0.496(5)$ is obtained by fitting the $w(N)$ data to the scaling $w(N) = w_0 N^{-b}$. The data for the same $N$ are represented by the same color in all panels. Error bars represent the standard error of the mean in all figures.

schematically in Fig. 2. The first level is formed by $N$ small networks, each of which takes as input a set of replica mean-squared displacements (rMSDs), $\{\Delta_{AB}^i\}$, where $i = 1 \ldots N$, and $A, B$ run over $M_i = N_i(N_i - 1)/2$ replica pairs (see Methods). The small network integrates the input data into a single output $q_i$, which is then fed into a big fully connected feedforward neural network (FNN) on the second level. The FNN predicts to which phase the input sample belongs – more specifically, it provides a probability $P(\hat{T}, N)$ (or $1 - P(\hat{T}, N)$) of an $N$-particle system belonging to phase I (or phase II), at a given temperature $\hat{T}$ (see Methods for the details.
of the machine learning methods) \cite{39,42}. In this study, we only perform binary phase classifications.

To determine the order of phase transition, a standard way is to apply a finite-size analysis of data obtained from experiments or simulations. For example, the fluctuation of order parameter, or the susceptibility, $\chi$, follows a finite-size scaling function around a phase transition,

$$\chi/N^a = \mathcal{X}\left(|\hat{T} - \hat{T}_c|N^b\right),$$

where $a, b$ are two exponents, $\mathcal{X}(x)$ a scaling function whose concrete form is not important to our discussion, and $\hat{T}_c$ the transition temperature. The values of $a$ and $b$ depend on the nature of transition: (i) For a standard first-order phase transition, $a = 1$ and $b = 1$ \cite{53}. An example is the first-order phase transition between positive and negative ferromagnetic phases in the Ising model under an external field. (ii) For a first-order phase transition in the presence of disorder, such as the yielding transition \cite{54} and the melting transition (see Fig. 3) in glasses, $a = 1$ and $b = 1/2$. Equation (1) then results in two susceptibilities, a disconnected one, $\chi_{\text{dis}} = \chi \sim N^a \sim N$, and a connected one $\chi_{\text{con}} \sim d\chi/d\hat{T} \sim N^b \sim N^{1/2}$. The two susceptibilities are related via, $\chi_{\text{dis}} \sim \chi_{\text{con}}^2$, a relation found in the canonical random field Ising model \cite{55,56}.

(iii) For a second-order phase transition, $a = b\gamma$ and $b = 1/d\nu$, where $d$ is the dimensionality, and $\gamma$ and $\nu$ are the critical exponents for the divergences of susceptibility and correlation length. A standard example is the second-order phase transition between paramagnetic and ferromagnetic phases in the Ising model without a field.

Similar finite-size analyses can be applied to the machine learning output \cite{39,42},

$$P(\hat{T}, N) = \mathcal{P}\left(|\hat{T} - \hat{T}_c|N^b\right),$$

where (i) $b = 1$ for a standard first-order phase transition, (ii) $b = 1/2$ for a first-order phase transition with disorder, and (iii) $b = 1/d\nu$ for a second-order phase transition. Note that the value of $b = 1/d\nu$ generally does not equal 1 or 1/2 (except for special cases such as the second-order phase transition in the 2D Ising model, see Supplementary Information (SI) Sec. S1). Thus by measuring $b$, we can distinguish between the abovementioned three kinds of phase transitions. Indeed, we find that (i) $b = 1$ for the first-order phase transition in the Ising model in both two and three dimensions (see SI Sec. S1), (ii) $b = 1/2$ for the glass melting, suggesting
that melting is a non-equilibrium first-order phase transition with disorder \( 35, 36 \), and (iii) \( b = 0.45(1) \) for the Gardner transition, suggesting that the Gardner transition is a non-equilibrium second-order phase transition, with a critical exponent \( \nu = 0.74(2) \), consistent with previous estimates \( 42, 43 \). Further tests are carried out for the equilibrium second-order phase transition in the Ising model in both two and three dimensions, confirming the relation \( b = 1/d\nu \) (see SI Sec. S1).

**Machine learning the melting transition**

Figure 3(a) shows the EOS, \( T(\varphi) \), obtained by MD simulations around the melting of ultra-stable glasses, together with the evolution of caging order parameter, \( \Delta(\varphi) \). The data show features similar to a discontinuous phase transition: the liquid and glass EOSs are connected by an abrupt jump, and the order parameter changes rapidly around melting. Interestingly, the behavior of the physically-defined caging order parameter \( \Delta \) is well reproduced, with proper rescaling, by the outcome \( q = \frac{1}{N} \sum_i q_i \) from the first-level neural networks, where \( \varphi \) represents a sample average. The agreement is further supported by the comparison between the distributions \( P(\Delta_i) \) and \( P(q_i) \) in both liquid and glass phases (see Fig. 3(b-c)), and by the strong correlations between \( \Delta_i \) and \( q_i \) (see Fig. 3(b-c) insets).

We next examine the finite-size effect around melting. The density susceptibility, \( \chi_\varphi = N \frac{\varphi^2 - (\bar{\varphi})^2}{(\bar{\varphi})^2} \), displays a clear dependence on system size \( N \), around the melting temperature \( T_m \approx 0.064 \) (see Fig. 4(a)). Its finite-size scaling satisfies Eq. (1), with \( a \approx 1 \) and \( b \approx 0.5 \) (see Fig. 4(b)). Consistently, the machine learning result \( P(\hat{T},N) \) agrees with Eq. (2), with \( b = 0.496(5) \) (see Fig. 4(c-d)). To determine \( b \), we first extract the width of transition regime, \( w(N) \), from the fitting of \( P(\hat{T},N) \), and then fit \( w(N) \) to a power-law scaling, \( w(N) \sim N^{-b} \) (see Fig. 4(d-inset)). Furthermore, as expected, the melting temperature \( T_m(N) \) obtained from machine learning data is around the peak position of \( \chi_\varphi(\hat{T}) \) (see Fig. 4(a) and (c-inset)).

**Machine learning the Gardner transition**

Compared to the melting transition, the Gardner transition has much more subtle signatures. Around the Gardner transition, the \( T(\varphi) \) EOS does not display any jump or kink \( 17 \). Practically, it is useful to examine the distribution \( P(\chi_i) \) of single-particle caging susceptibility, which shows a secondary peak in the marginally stable phase \( T < T_G \) \( 42 \). Interestingly, the features of \( P(\chi_i) \) can also be correctly reproduced by \( P(q_i) \) from the first-level networks, in both phases (see Fig. 5).

As expected, the finite-size scaling of \( P(\hat{T},N) \) output from the second-level network satisfies Eq. (2) as well (see Fig. 6). Based on the relation \( b = 1/d\nu \), we obtain an estimate of the critical exponent \( \nu = 0.74(2) \) (see Fig. 6(b-inset)), which is close to previous results \( \nu = 0.78(2) \) from a numerical analysis \( 42 \) and \( \nu = 0.85 \) from a field-theory calculation \( 43 \). The transition temperature \( T_G(N) \) (see Fig. 6(a-inset)) is also consistent with previous estimates \( 17, 42 \). Note that a similar machine learning approach has been applied in Ref. \( 42 \) to determine \( \nu \) and \( T_G(N) \). The key difference is that, Ref. \( 42 \) used the handcrafted \( \{\chi_i\} \) as the input to a single FNN, while here \{\chi_i\} can be autonomously extracted. Thus the current approach will be useful in more general situations when a deep, prior understanding of the underlying system is unavailable.

Because the Gardner transition is a second-order phase transition, we expect that the “ordered” phase \( T < T_G \) has two states. This is similar to the existence of two symmetric states (with positive and negative magnetizations) in the ferromagnetic phase of the Ising model below the critical temperature. To reveal the two states in the marginal glass phase, we apply unsupervised machine learning using the t-distributed stochastic neighbor embedding (t-SNE) method \( 57 \) (see Methods). Two groups of samples are identified (see Fig. 7(a)). The distributions of \( \chi_i \) of the two groups
correspond to the two peaks in $P(\chi_t)$ (see Fig. 7(b)), which validates the classification and confirms that the Gardner transition is a second-order phase transition.

**FIG. 8.** Learning the caging order parameter of the glass transition. (a) EOSs for $\hat{T}_g = 0.033$ (same as in Fig. 3(a)) and $\hat{T}_g = 0.047$ (for $N = 500$). The glass transition temperature $\hat{T}_g$ is practically defined as the intersection of the CS liquid EOS (green line) and the linearly fitted glass EOS (dashed line). (b) The average cage size $\Delta$ (measured with $t = 1$) as a function of $\varphi$, for ordinary glasses ($\hat{T}_g = 0.047$), together with the machine learned $\tilde{q}(\varphi)$ using different blanking windows $[T_2, T_1]$.

**Failure of learning the glass transition**

Naturally, one could ask whether the above method can be applied to the glass transition. For this purpose, we consider ordinary glasses compressed from low-density liquid states at an initial temperature $\hat{T} \approx 0.4$ ($\varphi \approx 0.2$) with a rate $\Gamma = 10^{-3}$. The corresponding glass transition occurs around $\hat{T}_g \approx 0.047$, much higher than $\hat{T}_g \approx 0.033$ of the above discussed ultra-stable states prepared by the swap algorithm (note that it is impossible to obtain ultra-stable glasses by MD compression). The difference between the two kinds of glasses, ordinary and ultra-stable, is revealed by the hysteresis in their EOSs (Fig. 8(a)). For an additional comparison on their dynamics, such as the mean-squared displacements, please see SI Sec. S4 for ordinary glasses and Ref. [17] for ultra-stable glasses.

Around $\hat{T}_g$, the average cage size $\Delta$ of ordinary glasses depends strongly on the measurement time $t$ (see Fig. S13). Nevertheless, for a fixed $t$, the evolution $\Delta(\varphi)$ can still be accurately captured by $\tilde{q}$ from machine learning (see Fig. 8(b)). However, the machine learning algorithm fails to provide a unique “transition point” (see Fig. 9). The estimated crossover from liquids to glasses, $\hat{T}^\text{ML}_g$, relies on the choice of algorithm parameters to define a blanking window (see Methods for the definition). Fig. 9(b) shows that $\hat{T}^\text{ML}_g$ is nearly identical to the center of blanking window, $\hat{T}_\text{center}$, suggesting that $\hat{T}^\text{ML}_g$ simply separates low-density and high-density states, whose definitions are preset artificially by the blanking window. Note that a real phase transition point should be independent of any algorithm parameters. The tests of melting and Gardner transitions confirm this point (see Figs. S8 and S10). The failure of our attempt to learn the glass transition $\hat{T}_g$ demonstrates that, either liquids and glasses do not have essential differences, or their characteristics are too delicate to be
learned by our current algorithm. Indeed, it is widely accepted that glass transition is not a well-defined phase transition, since one cannot observe any divergence or discontinuity at \( T^* \).

Discussion

Glasses are out of equilibrium, and as a consequence, all three transitions discussed here should, in principle, depend on the time scale of interest. The finite-time effect of the Gardner transition in ultra-stable glasses is discussed in detail in Refs. [17, 42]. In SI Sec. S2, we also show how the melting transition depends on the decomposition rate. Our strategy here is to fix firstly a time scale \( t \) and then examine the finite-size effect as for standard equilibrium phase transitions. As shown by this study and [42], such an approach is applicable for stable glasses, where the simulation time \( t \) and the structural relaxation time \( \tau_n \) are well separated. This precondition, however, breaks down for ordinary glasses. Thus the transition from liquids to ordinary glasses has very different properties from standard equilibrium phase transitions. As a result, an ordinary glass transition could not be properly machine learned.

To conclude, we propose to machine learn dynamical, instead of structural, features of liquids and glasses. The caging dynamics are directly related to the structure of phase space. In the language of replica theory, the breakdown of ergodicity at the glass transition is a result of replica symmetry breaking (RSB) [27]. The two-level structure of NNN can naturally capture caging order parameters of both the melting transition (1-step RSB) and the Gardner transition (full-step RSB). Increasing the number of nested levels would be useful for other types of RSB (e.g., 2-step RSB in glasses of bidisperse particles [58]). Finally, we expect generalizations of our method to other glassy systems, such as polymers and spin glasses.

Methods

Glass model

The model [17, 34, 47, 59, 60] consists of \( N = 125 - 8000 \) polydisperse hard spheres, whose diameters are distributed according to a continuous function \( P_D(D_{\text{min}} \leq D \leq D_{\text{min}}/0.45) \sim D^{-3} \). The volume of simulation box is \( V \), and periodic boundary conditions are used. The system state is characterized by volume fraction \( \varphi \) and reduced temperature \( T = 1/P = Nk_B T/ \hat{P} V \), where \( P \) is the pressure, \( \hat{P} \) the reduced pressure, \( k_B = 1 \) the Boltzmann constant, and \( T = 1 \) the temperature. We set the mean diameter \( D_{\text{mean}} \) as unit length, and the particle mass \( m \) as unit mass. Crystallization is suppressed by polydispersity, and will not be discussed in this study.

Molecular dynamics simulation

We use the Lubachevsky-Stillinger algorithm (event-driven MD) to simulate compression quench [61, 62]. During compression/decompression, the sizes of all particles are increased/decreased proportionally with a fixed rate \( \Gamma = \frac{1}{\sqrt{m D_{\text{mean}}^2/ \left( k_B T \right)}} \). The simulation time is expressed in units of \( \sqrt{m D_{\text{mean}}^2/ \left( k_B T \right)} \).

Swap algorithm

The swap algorithm simulates artificial dynamics that can efficiently accelerate reaching equilibrium [34]. At each swap Monte Carlo (MC) step, two particles are randomly picked, and swapped if they do not overlap with neighbor particles at the new positions. While the dynamics are unrealistic, the final configurations in equilibrium are equivalent to those generated by standard MD and MC simulations.

Caging order parameters

Let us consider \( N_r \) replicas of the same sample. According to the replica theory [27], a replicated configuration \( C \) is drawn from a given distribution \( P(C) \) depending on the state of the system. In liquids, \( P(C) \) is the equilibrium distribution and \( C \) can be any microscopic state in the ensemble. In glasses, there are multiple meta-stable states \( C_{\text{MS}} \) and different samples may belong to different meta-stable states. Thus the replicas of a given sample should be constructed from the conditional probability distribution \( P(C|C_{\text{MS}}) \), where \( C_{\text{MS}} \) is the meta-stable state that the given sample belongs to. From a dynamical point of view, it means that the same particle from different replicas should be confined in the same cage, but it is allowed to vibrate freely inside the cage. The purpose of replica construction is to translate caging dynamics to a static distribution of replicas. With this setup, the properties of cages can be obtained by analyzing the ensemble of replicas.

There are many ways to quantify caging order, such as the average cage size or the fluctuation (susceptibility) of cage sizes. Below are three quantities that are relevant to this study. The main idea is to compute pairwise quantities for two replicas \( A \) and \( B \) randomly drawn from the ensemble of replicas.

Single-particle cage size

From an ensemble of \( N_r \) replicas, one can make in total \( M_r = N_r (N_r - 1)/2 \) pairs of replicas. The rMSD is [17, 63].

\[
\Delta_{AB}^i = \left| \mathbf{r}_i^A - \mathbf{r}_i^B \right|^2,
\]

where \( \mathbf{r}_i^A \) and \( \mathbf{r}_i^B \) are the positions of particle \( i \) in replicas \( A \) and \( B \). The single-particle cage size of particle \( i \) is

\[
\Delta_i = \langle \Delta_{AB}^i \rangle = \frac{2}{N_r (N_r - 1)} \sum_{A \neq B} \Delta_{AB}^i,
\]

where \( \langle x \rangle \) represents the average over replica pairs.

Average cage size
The cage size averaged over both particles and samples (quenched disorder) is,
\[
\Delta = \frac{1}{N} \sum_{i=1}^{N} \Delta_i, \tag{5}
\]

**Single-particle caging susceptibility**

The single-particle caging susceptibility is defined as
\[
\chi_i = \langle (u_{AB}^i)^2 \rangle - \langle u_{AB}^i \rangle^2 = \frac{\langle (\Delta_{AB}^i)^2 \rangle - \langle \Delta_{AB}^i \rangle^2}{\langle \Delta_{AB}^i \rangle^2}, \tag{6}
\]
where
\[
u_{AB}^i = \frac{\Delta_{AB}^i - \langle \Delta_{AB}^i \rangle}{\langle \Delta_{AB}^i \rangle}.
\tag{7}
\]

It can be seen that by definition, \( \langle u_{AB}^i \rangle = 0 \).

So far we have provided mathematical definitions of caging order parameters. In **Input data to machine learning**, we will explain in detail how to construct replicas from MD simulations.

**Architecture of the artificial neural network**

The NNN comprises two levels of networks, which in general can have different structures (see Fig. 2). There are \( N \) duplicated small networks at the first level, each of which extracts the latent caging features of one single particle. The small network has only one hidden layer, besides the input and output layers. Both input and hidden layers have \( M_r \) neuron nodes, and the output layer has a single node. The \( i^{\text{th}} \) hidden node is connected by a single link to the \( i^{\text{th}} \) input node, and is activated by the exponential linear unit (ELU) function. The output node simply takes an average of \( M_r \) hidden nodes. The \( N \) small networks share the same parameters (weights and bias), and thus there are only \( 2M_r \) free parameters at the first level. In principle, one could choose other architectures (e.g., FNN) for small networks, and set their parameters to be independent. In practice, however, we find that using a small number of free parameters at the first level can significantly increase the efficiency of the NNN model during training, without losing its compatibility and predictive power.

The \( N \) output nodes of the first-level small networks are considered as input nodes of the followed big FNN at the second-level. The FNN has one hidden layer of 128 nodes activated by ELU functions, and one output layer of two nodes that provide binary classifications through softmax activation functions.

**Input data to machine learning**

For the system of \( N \) particles at a temperature of \( T \), we perform molecular simulations to obtain \( N_{s}^{\text{train}} \) independent samples, and divide them into three non-overlapping data sets for the purposes of training \( (N_{s}^{\text{train}} \) samples), validation \( (N_{s}^{\text{valid}} \) samples) and prediction \( (N_{s}^{\text{pred}} \) samples), where \( N_{s}^{\text{total}} = N_{s}^{\text{train}} + N_{s}^{\text{valid}} + N_{s}^{\text{pred}} \).

Below we explain in detail how the input data to NNN are prepared, for the three types of transitions, accordingly.

**Glass transition**

The initial state of each sample is a dilute liquid configuration at \( \varphi = 0.2 \). Each sample is compressed to a target density \( \varphi \) (alternatively one can set a target temperature \( \hat{T} \), or pressure \( \hat{P} \)) with a fixed rate \( \Gamma \), using the Lubachevsky-Stillinger algorithm. We choose in total \( N_{T} \) different target densities in a window \( \hat{T} \in (0.002, 0.268) \). The system is in a liquid state if \( \varphi < \varphi_{g} \); otherwise in a glass state. Once the target \( \varphi \) is reached, we stop compression and make \( N_r \) replicas of each sample. These replicas share the same particle positions, but the particle velocities are independently reset according to the Maxwell–Boltzmann distribution. After making \( N_r \) replicas, we reset simulation time \( t \) to zero, and then perform constant volume (\( \Gamma = 0 \)) MD simulations. Note that the \( N_r \) replicas evolve independently because they are assigned to different velocities at \( t = 0 \). We then collect configurations of these replicas at \( t > 0 \), and compute the rMSD, \( \Delta_{AB}^i \), as defined in Eq. (3). For the \( N_r \) replicas, there are in total \( M_{t} = N_{r}(N_{r} - 1)/2 \) pairs can be formed, \( \{A, B\} = \{(1, 1), (1, 2), \ldots, (N_{r} - 1, N_{r})\} \). The vector of \( \{\Delta_{AB}^i\} = \{\Delta_{11}^i, \Delta_{12}^i, \ldots, \Delta_{N_{r}-1,N_{r}}^i\} \) is fed into the \( i^{\text{th}} \) small network at the first level (see Fig. 2). The complete input data is a vector of \( N \times M_{t} \) element, \( \{\{\Delta_{1A}^i\}, \{\Delta_{2A}^i\}, \ldots, \{\Delta_{N_{r}A}^i\}\} \). For the data presented in this study, we have used \( \Gamma = 10^{-3} \), \( N_{s}^{\text{total}} = 1152 \), \( N_{s}^{\text{train}} = 864 \), \( N_{s}^{\text{valid}} = 144 \), \( N_{s}^{\text{pred}} = 144 \), and \( N_T = 29 \).

**Melting transition**

The initial state of each sample is an equilibrium liquid configuration at \( T_{g} = 0.033 \), generated by the swap algorithm. We then make \( N_r \) replicas of each sample, and decompress them using the Lubachevsky-Stillinger algorithm with a negative \( \Gamma = -10^4 \), to a series of target temperatures \( \hat{T} \) in a window \( \hat{T} \in (0.033, 0.123) \). After decompression, the rMSD, \( \Delta_{AB}^i \), are computed and used as the input to NNN. Note that this procedure is different from the one explained above for the glass transition: the replicas are created before decompression/compression since the ultra-stable configurations prepared by the swap algorithm are already in the glass phase. We set \( N_{s}^{\text{total}} = 2400 \), \( N_{s}^{\text{train}} = 1800 \), \( N_{s}^{\text{valid}} = 300 \), \( N_{s}^{\text{pred}} = 300 \), and \( N_T = 35 \).

**Gardner transition**

The procedure is similar to the one for the melting transition, except that the compression rate \( \Gamma = 10^{-4} \) is positive. Instead of \( \{\Delta_{AB}^i\} \), we use the normalized quantity \( \{u_{AB}^i\} \) (see Eq. (7)) as the input data to NNN. It turns out that such a simple pre-treatment can efficiently improve the performance of our machine learning.
model (see SI. Sec. S3A). The following parameters are used: \( N_{\text{total}}^{s} = 2400, N_{s}^{\text{train}} = 1800, N_{s}^{\text{valid}} = 300, N_{s}^{\text{pred}} = 300, \) and \( N_{P} = 23 \) (target \( \hat{T} \) is chosen in a window \( \hat{T} \in (0, 0.033) \)).

Blanking window

For the supervised learning of phases, we need to label in advance to which phase a given configuration belongs, during training and validation. A blanking window \( [T_2, T_1] \) is introduced to skip the vicinity of a (presumed) transition. Specifically, the following setup is used for the data presented in the main figures: for the melting transition, configurations at \( \hat{T} > \hat{T}_1 = 0.083 \) and \( \hat{T} < \hat{T}_2 = 0.053 \) are labeled as liquids and glasses respectively; for the Gardner transition, configurations at \( \hat{T} > \hat{T}_1 = 0.011 \) and \( \hat{T} < \hat{T}_2 = 0.0045 \) are labeled as stable and marginal glasses respectively. Note that, for both transitions, the machine learning results do not sensitively depend on the choice of blanking window (see SI. Secs. S2C and S3B). In contrast, the learning results of the glass transition correlates strongly to the blanking window (see Fig. [9]).

Training NNN and validating predictions

The cross-entropy cost function is minimized during training. The Adam optimizer \([64]\) is used to implement the stochastic gradient descent method for updating the network parameters. To avoid overfitting, a dropout strategy \([65]\) is used, which randomly skips 20% hidden nodes at each step. To augment the training data set, we perform \( N_{\text{shuffle}} = 20 \sim 200 \) random shuffles of the elements in the input vector, which is equivalent to randomly ordering particle indexes. In this way, we expand the training data set to \( N_{s}^{\text{train}} \times N_{P} \times N_{\text{shuffle}} \sim 10^5 \) samples. The random shuffling apparently destroys spatial correlations (if there is any) between particles. However, we find that it does not modify the final predictions noticeably. Validation is performed after each training step, by calculating the cost function for the validation data set. The entire training procedure is terminated when the validation cost function reaches a minimum. Such an early stopping strategy can efficiently avoid overfitting.

Making predictions using NNN

Once well trained, the NNN can make phase predictions for the samples in the testing data set. For each test sample at a temperature \( \hat{T} \), the NNN provides an output value of 0 or 1. The arithmetic mean of the output over all test samples gives an estimation of the probability \( P \) of the system belonging to a specific phase, and \( 1 \sim P \) to the other.

To achieve a reliable prediction, we independently train our NNN for 10 times (runs) and calculate the mean and the statistical error of the predicted \( P \). For each run, \( N_{s}^{\text{train}} \) training samples and \( N_{s}^{\text{pred}} \) prediction samples are randomly selected from \( N_{s}^{\text{total}} \) samples, and the remaining \( N_{s}^{\text{valid}} \) samples are used for validation.

Unsupervised classification using t-distributed stochastic neighbor embedding

We utilize the unsupervised t-distributed stochastic neighbour embedding (t-SNE) method \([57]\) to group samples in the Gardner phase. The input data are single-particle caging susceptibilities \( \{\chi_i\} \) of each sample, where \( i = 1, 2, \ldots, N \). The algorithm conducts a nonlinear dimensionality reduction, which maps each vector \( \{\chi_i\} \) to a point in two dimensional space, data points are rearranged according to their similarities quantified by a t-distribution kernel function.

Acknowledgments

Y. Jin acknowledges funding from Project 11974361, Project 11935002 and Project 12047503 supported by National Natural Science Foundation of China, from the Key Research Program of Frontier Sciences, Chinese Academy of Sciences, Grant NO. ZDBS-LY-7017, and from the Key Research Program of the Chinese Academy of Sciences, Grant NO. XDPB15. Y. Jiang acknowledges funding from Project 22073004 supported by National Natural Science Foundation of China. This work was granted access to the High-Performance Computing Cluster of Institute of Theoretical Physics - Chinese Academy of Sciences and Beihang University.

Additional Information

Data availability

The data are available from the corresponding authors on request.

Author contributions

All authors contributed equally to the paper.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Correspondence and material requests should be addressed to Y. Jin (yuliangjin@mail.ita.ac.cn) or Y. Jiang (yjiang@buaa.edu.cn).

[1] C Austen Angell, “Formation of glasses from liquids and biopolymers,” Science 267, 1924–1935 (1995).
[2] C Patrick Royall and Stephen R Williams, “The role of local structure in dynamical arrest,” Physics Reports 560, 1–75 (2015).
[3] Hajime Tanaka, Hua Tong, Rui Shi, and John Russo, “Revealing key structural features hidden in liquids and glasses,” Nature Reviews Physics 1, 333–348 (2019).
[4] Paul J Steinhardt, David R Nelson, and Marco Ronchetti, “Bond-orientational order in liquids and glasses,” Physical Review B 28, 784 (1983).
[5] Burkhard Doliwa and Andreas Heuer, “What does the
potential energy landscape tell us about the dynamics of supercooled liquids and glasses?” Physical Review Letters 91, 235501 (2003).

[6] Ludovic Berthier and Robert L Jack, “Structure and dynamics of glass formers: Predictability at large length scales,” Physical Review E 76, 041509 (2007).

[7] Asaph Widmer-Cooper, Heidi Perry, Peter Harrowell, and David R Reichman, “Irreversible reorganization in a supercooled liquid originates from localized soft modes,” Nature Physics 4, 711–715 (2008).

[8] Chengjie Xia, Jindong Li, Yixin Cao, Binquan Kou, Xianghui Xiao, Kamel Fezzaa, Tiqiao Xiao, and Yujie Wang, “The structural origin of the hard-sphere glass transition in granular packing,” Nature Communications 6, 1–9 (2015).

[9] Susana Marín-Aguilar, Henricus H Wensink, Giuseppe Foffi, and Frank Smol, “Tetrahedrality dictates dynamics in hard sphere mixtures,” Physical Review Letters 124, 208005 (2020).

[10] Mathieu Leocmach and Hajime Tanaka, “Roles of icosahedral and crystal-like order in the hard spheres glass transition,” Nature Communications 3, 1–8 (2012).

[11] Akihiko Hirata, Pengfei Guan, Takeshi Fujita, Yoshihiko Hirotsu, Akhisa Inoue, Alain Reza Yavari, Toshio Sasaki, and Mingwei Chen, “Direct observation of local atomic order in a metallic glass,” Nature Materials 10, 28–33 (2011).

[12] YC Hu, FX Li, MZ Li, HY Bai, and WH Wang, “Fivefold symmetry as indicator of dynamic arrest in metal glass-forming liquids,” Nature Communications 6, 1–8 (2015).

[13] Jade Taffs and C Patrick Royall, “The role of fivefold symmetry in suppressing crystallization,” Nature Communications 7, 1–7 (2016).

[14] Hua Tong and Hajime Tanaka, “Revealing hidden structural order controlling both fast and slow glassy dynamics in supercooled liquids,” Physical Review X 8, 011041 (2018).

[15] Hua Tong and Hajime Tanaka, “Structural order as a genuine control parameter of dynamics in simple glass formers,” Nature Communications 10, 1–10 (2019).

[16] Daniel B Miracle, “A structural model for metallic glasses,” Nature Materials 3, 697–702 (2004).

[17] Ludovic Berthier, Patrick Charbonneau, Yuliang Jin, Giorgio Parisi, Beatriz Seoane, and Francesco Zamponi, “Growing timescales and length scales characterizing vibrations of amorphous solids,” Proceedings of the National Academy of Sciences 113, 8397–8401 (2016).

[18] Ekin D Cubuk, Samuel Stern Schoenholz, Jennifer M Rieser, Brad Dean Malone, Joerg Rottler, Douglas J Durian, Efthimiios Kaxiras, and Andrea J Liu, “Identifying structural flow defects in disordered solids using machine-learning methods,” Physical Review Letters 114, 108001 (2015).

[19] Samuel S Schoenholz, Ekin D Cubuk, Daniel M Sussman, Efthimiios Kaxiras, and Andrea J Liu, “A structural approach to relaxation in glassy liquids,” Nature Physics 12, 469–471 (2016).

[20] Ekin Dogus Cubuk, RJS Ivancic, Samuel Schoenholz, DJ Strickland, Anindita Basu, ZS Davidson, Julien Fontaine, Jyo Lyn Hor, Y-R Huang, Y Jiang, et al., “Structure-property relationships from universal signatures of plasticity in disordered solids,” Science 358, 1033–1037 (2017).

[21] Samuel S Schoenholz, Ekin D Cubuk, Efthimiios Kaxiras, and Andrea J Liu, “Relationship between local structure and relaxation in out-of-equilibrium glassy systems,” Proceedings of the National Academy of Sciences 114, 263–267 (2017).

[22] Victor Bapst, Thomas Keck, A Grabska-Barwińska, Craig Donner, Ekin Dogus Cubuk, Samuel S Schoenholz, Annette Obika, Alexander WR Nelson, Trevor Back, Denis Hassabis, et al., “Unveiling the predictive power of static structure in glassy systems,” Nature Physics 16, 448–454 (2020).

[23] Emanuele Boattini, Frank Smol, and Laura Filion, “Averaging local structure to predict the dynamic propensity in supercooled liquids,” Phys. Rev. Lett. 127, 088007 (2021).

[24] Joris Paret, Robert L Jack, and Daniele Coslovich, “Assessing the structural heterogeneity of supercooled liquids through community inference,” The Journal of Chemical Physics 152, 144502 (2020).

[25] Emanuele Boattini, Susana Marín-Aguilar, Saheli Mitra, Giuseppe Foffi, Frank Smol, and Laura Filion, “Autonomously revealing hidden local structures in supercooled liquids,” Nature Communications 11, 1–9 (2020).

[26] Giorgio Parisi and Francesco Zamponi, “Mean-field theory of hard glasses and jamming,” Reviews of Modern Physics 82, 759 (2010).

[27] Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi, Theory of Simple Glasses: Exact Solutions in Infinite Dimensions (Cambridge University Press, 2020).

[28] Marc Mézard, Giorgio Parisi, and Miguel Virasoro, Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications, Vol. 9 (World Scientific Publishing Company, 1987).

[29] Wolfgang Götte, Complex dynamics of glass-forming liquids: A mode-coupling theory, Vol. 143 (OUP Oxford, 2008).

[30] David R Reichman and Patrick Charbonneau, “Mode-coupling theory,” Journal of Statistical Mechanics: Theory and Experiment 2005, P05013 (2005).

[31] Liesbeth Janssen, “Mode-coupling theory of the glass transition: A primer,” Frontiers in Physics 6, 97 (2018).

[32] TR Kirkpatrick and PG Wolynes, “Connections between some kinetic and equilibrium theories of the glass transition,” Physical Review A 35, 3072 (1987).

[33] Theodore R Kirkpatrick, Devarajan Thirumalai, and Peter G Wolynes, “Scaling concepts for the dynamics of viscous liquids near an ideal glassy state,” Physical Review A 40, 1045 (1989).

[34] Ludovic Berthier, Daniele Coslovich, Andrea Ninarello, and Misaki Ozawa, “Equilibrium sampling of hard spheres up to the jamming density and beyond,” Physical Review Letters 116, 238002 (2016).

[35] Robert L Jack and Ludovic Berthier, “The melting of stable glasses is governed by nucleation-and-growth dynamics,” The Journal of Chemical Physics 144, 244506 (2016).

[36] Ludovic Berthier and Robert L Jack, “Evidence for a disordered critical point in a glass-forming liquid,” Physical Review Letters 114, 205701 (2015).

[37] Elisabeth Gardner, “Spin glasses with p-spin interactions,” Nuclear Physics B 257, 747–765 (1985).

[38] Patrick Charbonneau, Jorge Kurchan, Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi, “Frac-
tial free energy landscapes in structural glasses,” Nature Communications 5, 3725 (2014).

[39] Juan Carrasquilla and Roger G. Melko, “Machine learning phases of matter,” Nature Physics 13, 431 (2017).

[40] Dongkyu Kim and Dong-Hee Kim, “Emergence of a finite-size-scaling function in the supervised learning of the ising phase transition,” Journal of Statistical Mechanics: Theory and Experiment 2021, 023202 (2021).

[41] Humberto Munoz-Bauza, Firas Hamze, and Helmut G Katzgraber, “Learning to find order in disorder,” Journal of Statistical Mechanics: Theory and Experiment 2020, 073302 (2020).

[42] Huaping Li, Yuliang Jin, Ying Jiang, and Jeff ZY Chen, “Determining the nonequilibrium criticality of a gardner transition via a hybrid study of molecular simulations and machine learning,” Proceedings of the National Academy of Sciences 118 (2021).

[43] Patrick Charbonneau and Sho Yaida, “Nontrivial critical fixed point for replica-symmetry-breaking transitions,” Physical Review Letters 118, 215701 (2017).

[44] Tomáš Boublík, “Hard-sphere equation of state,” The Journal of Chemical Physics 53, 471–472 (1970).

[45] Sadanand Singh, Mark D Ediger, and Juan J De Pablo, “Ultrastable glasses from in silico vapour deposition,” Nature Materials 12, 139–144 (2013).

[46] Stephen F Swallen, Katherine Traynor, Robert J McMahon, MD Ediger, and Thomas E Mates, “Stable glass transformation to supercooled liquid via surface-initiated growth front,” Physical Review Letters 102, 065503 (2009).

[47] Yuliang Jin and Hajime Yoshiba, “Exploring the complex free-energy landscape of the simplest glass by rheology,” Nature Communications 8, 14935 (2017).

[48] Qinyi Liao and Ludovic Berthier, “Hierarchical landscape of hard disk glasses,” Physical Review X 9, 011049 (2019).

[49] Antoine Seguin and Olivier Dauchot, “Experimental evidence of the gardner phase in a granular glass,” Physical Review Letters 117, 228001 (2016).

[50] Korbinian Geirhos, Peter Lunkenheimer, and Alois Loidl, “Johari-goldstein relaxation far below t g: Experimental evidence for the gardner transition in structural glasses?,” Physical Review Letters 120, 085705 (2018).

[51] Andrew P. Hammond and Eric I. Corwin, “Experimental observation of the marginal glass phase in a colloidal glass,” Proceedings of the National Academy of Sciences 117, 5714–5718 (2020).

[52] Patrick Charbonneau, Jorge Kurchan, Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi, “Glass and jamming transitions: From exact results to finite-dimensional descriptions,” Annual Review of Condensed Matter Physics 8, 265–288 (2017).

[53] Kurt Binder, “Theory of first-order phase transitions,” Reports on progress in physics 50, 783 (1987).

[54] Misaki Ozawa, Ludovic Berthier, Giulio Biroli, Alberto Rosso, and Gilles Tarjus, “Random critical point separates brittle and ductile yielding transitions in amorphous materials,” Proceedings of the National Academy of Sciences 115, 6656–6661 (2018).

[55] Michael Gofman, Joan Adler, Amnon Aharony, A Brooks Harris, and Moshe Schwartz, “Evidence for two exponent scaling in the random field ising model,” Physical Review Letters 71, 1569 (1993).

[56] Thomas Nattermann, “Theory of the random field ising model,” in Spin glasses and random fields (World Scientific, 1998) pp. 277–298.

[57] Laurens van der Maaten and Geoffrey Hinton, “Visualizing data using t-sne,” Journal of Machine Learning Research 9, 2579–2605 (2008).

[58] Harukuni Ikeda, Kunimasu Miyazaki, Hajime Yoshiba, and Atushi Ikeda, “Multiple glass transitions and higher-order replica symmetry breaking of binary mixtures,” Physical Review E 103, 022613 (2021).

[59] Yuliang Jin, Pierfrancesco Urbani, Francesco Zamponi, and Hajime Yoshiba, “A stability-reversibility map unifies elasticity, plasticity, yielding, and jamming in hard sphere glasses,” Science Advances 4, eaat6387 (2018).

[60] Beatriz Seoane and Francesco Zamponi, “Spin-glass-like aging in colloidal and granular glasses,” Soft Matter 14, 5222–5234 (2018).

[61] Boris D Lubachevsky and Frank H Stillinger, “Geometric properties of random disk packings,” Journal of Statistical Physics 60, 561–583 (1990).

[62] Monica Skoge, Aleksandar Donev, Frank H Stillinger, and Salvatore Torquato, “Packing hyperspheres in high-dimensional euclidean spaces,” Physical Review E 74, 041127 (2006).

[63] Patrick Charbonneau, Yuliang Jin, Giorgio Parisi, Corrado Ramone, Beatriz Seoane, and Francesco Zamponi, “Numerical detection of the gardner transition in a mean-field glass former,” Physical Review E 92, 012316 (2015).

[64] Diederik P. Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” International Conference on Learning Representations (2015).

[65] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” The journal of Machine Learning Research 15, 1929–1958 (2014).

[66] Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al., “Tensorflow: A system for large-scale machine learning,” in 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16) (2016) pp. 265–283.

[67] M. E. J. Newman and G. T. Barkema, Monte Carlo methods in statistical physics (Clarendon Press, Oxford, 1999).

[68] Lars Onsager, “Crystal statistics. i. the simple-cubic lattice,” Physical Review E 7, 1566 (1944).

[69] Michael E Fisher and Robert J Burford, “Theory of critical-point scattering and correlations. i. the ising model,” Physical Review 156, 583 (1967).

[70] AL Talapov and HWJ Blöte, “The magnetization of the ising phase transition,” Journal of Statistical Mechanics: Theory and Experiment 9, 2579–2605 (2008).

[71] Massimo Campostrini, Andrea Pelissetto, Paolo Rossi, AL Talapov and HWJ Blöte, “The magnetization of the ising model,” in Spin glasses and random fields (World Scientific, 1998) pp. 277–298.

[72] Rui Zhang, Bin Wei, Dong Zhang, Jia-Ji Zhu, and Kai Chang, “Few-shot machine learning in the three-dimensional ising model,” Physical Review B 99, 094427 (2019).
Supplementary Information

S1. MACHINE LEARNING PHASE TRANSITIONS IN THE ISING MODEL

A. Machine learning algorithm

We study the Ising model in both two and three dimensions to validate the finite-size scaling function Eq. (2) in a standard equilibrium system. Machine learning is carried out using TensorFlow [66]. Following [39], we make use of a fully connected feed-forward neural network (FNN), which comprises three layers (input, hidden and output) of nodes. The number of neurons in the input layer is equal to the number of spins $N = L^d$, where $L$ is the linear size of the lattice and $d$ is the dimensionality. The hidden layer is composed of 200 neurons activated by sigmoid functions, and the output layer has 2 neurons activated by a softmax function. During training, a cross-entropy cost function is minimized by means of a stochastic gradient descent method with an Adam optimizer [64]. In order to avoid overfitting, we adopt a dropout regularization [65] and an early stopping strategy.

The input data are spin configurations generated by the Wolff algorithm [67]. At a given magnetic field $H$ and a given temperature $T$, we prepare $N_t \approx 10,000$ samples, and use 80% of them for training, 10% for validation and 10% for prediction. Random shuffling is applied two to four times to make sure that there are sufficient samples at each combination of $(H,T)$ during training.

B. Learning the second-order phase transition

In two and three dimensions, a second order phase transition occurs at $T_c$ when the temperature is varied under the zero-field condition $H = 0$. Previous studies have established the values of $T_c$ and $\nu$ (the critical exponent for the divergence of correlation length): $T_c \approx 2.26918531421 \ldots$ [68] and $\nu = 1$ [69] in 2D; $T_c \approx 4.511528(6)$ [70] and $\nu \approx 0.63012(16)$ [71] in 3D. Supervised machine learning techniques have been well utilized to learn the continuous phase transition in the Ising model in both 2D [39] and 3D [72]. Here we reproduce these results using our algorithm. For this pur-
pose, we generate zero-field ($H = 0$) input configurations around $T_c$ at $N_T$ different temperature points. The data points $P(T, N)$ obtained from machine learning are fitted to

$$P(T, N) = \frac{1}{2} + \frac{1}{2} \text{erf}\{[T - T_c(N)]/w(N)\}, \quad (S1)$$

where $N = L^d$, erf($x$) is the error function, and $T_c(N)$ and $w(N)$ are two fitting parameters representing the critical temperature and the width of transition region (see Figs. S1 for 2D and Fig. S2 for 3D). The estimated critical temperatures agree with existing values (see insets of Figs. S1 and S2). Next, we examine the finite-size scaling function Eq. (2). The scaling function suggests that,

$$w(N) \sim N^{-b}, \quad (S2)$$

which is used to determine the critical exponent, $\nu = 0.967(4)$ (or $b = 1/\nu = 0.517(4)$) in 2D and $\nu = 0.631(3)$ (or $b = 0.528(4)$) in 3D (see Fig. S5). These estimations are in a good agreement with the standard values, $\nu = 1$ in 2D [69] and $\nu \approx 0.63012$ in 3D [71]. Indeed, the data points of $P(T, L)$ for different sizes can be collapsed onto a single curve if they are plotted as a function of the rescaled quantity $(T - T_c)N^{1/d\nu}$ (see Figs. S1(b) and S2(b)).

C. Learning the first-order phase transition

For a fixed temperature $T < T_c$, a first-order phase transition occurs at $H_c = 0$ when $H$ is varied. To our knowledge, the finite-size scaling Eq. (2) of the first-order phase transition in the Ising model has not been systematically studied yet within the machine learning framework. In this study, we show that the expected scaling Eq. (2) is fully consistent with our data. We set $k_B T/J = 2.1$ for the 2D model and $k_B T/J = 4.0$ for the 3D model, where $J$ is the interaction constant. Configurations are generated at $N_H$ different external fields around $H_c = 0$, with positive and negative fields evenly divided. The data points $P(H, N)$ obtained from machine learning are fitted to

$$P(H, N) = \frac{1}{2} + \frac{1}{2} \text{erf}\{[H - H_c(N)]/w(N)\}. \quad (S3)$$

As shown in Figs. S3 and S4, the estimated transition field is close to $H_c = 0$. Furthermore, we obtain $b \approx 0.96(2)$ for the 2D model, and $b \approx 1.01(4)$ for the 3D model.
FIG. S5. Comparing the finite size scalings of first- and second-order phase transitions in the Ising model. The data of transition width $w(N)$ are obtained from Figs. S1-S4. The exponent $b$ is obtained from fitting $w(N) = w_0 N^{-b}$ (lines). (a) In the 2D Ising model, we obtain $b = 0.96(2)$ for the first-order phase transition, and $b = 1/d\nu = 0.517(4)$ (i.e., $\nu = 0.967(4)$) for the second-order phase transition. (b) In the 3D Ising model, we obtain $b = 1.012(4)$ for the first-order phase transition, and $b = 0.528(4)$ (i.e., $\nu = 0.631(3)$) for the second-order phase transition.

model, which are consistent with the expected value $b = 1$ [53].

D. Distinguishing between first- and second-order phase transitions

Based on above analyses, we confirm that, by utilizing the scaling function Eq. (2), the original machine learning approach proposed in [39] can be generalized to identify both first- and second-order phase transitions, at least in the standard Ising model. Very importantly, the order of phase transition can be identified because the finite-size exponents $b$ in Eq. (2) are distinguishable within the numerical accuracy for first- and second-order phase transitions. As shown in Fig. S5, $b = 1$ for first-order phase transitions (without considering the effect of disorder), and $b = 1/d\nu$ for second-order phase transitions. While the phase transitions in the Ising model are in equilibrium, we show that the approach can be further generalize to non-equilibrium first-order (melting transition) and second-order (Gardner transition) phase transitions in disordered systems such as glasses (see the main text and Fig. S11 for details).

S2. ADDITIONAL DATA FOR THE MELTING TRANSITION

A. Dependence of the melting transition on the decompression rate

The $\hat{T} - \varphi$ equations of states (EOSs) of ultra-stable glasses in Fig. S6 show that the melting transition temperature $T_m$ decreases with slower decompression. It is expected that, in the limit $\Gamma \to 0$, the hysteresis in EOS will disappear and the glass melting will become a continuous crossover. However, for the range of $\Gamma$ relevant to this study, the discontinuous feature remains. In the main text, we examine the finite-size effect for a fixed decompression rate $\Gamma = -10^{-4}$, and do not further discuss the rate-dependence.

FIG. S6. Evolution of temperature $\hat{T}$ as a function of volume fraction $\varphi$ under decompression, for a few different decompression rate $\Gamma$. The system is composed of $N = 125$ particles, and is decompressed from ultra-stable states at $\{\varphi_g = 0.63, T_g = 0.033\}$. The liquid and glass EOSs are connected around the melting temperature $T_m$, which decreases with slower decompression.
example. We find that the values of \( c_1 \) and \( c_2 \) are non-deterministic, but the agreement between the rescaled predictions and the physically defined caging order parameters is very robust (Fig. S7(b)). It implies that the evolution of caging order parameter is correctly captured by NNN, but it is unnecessarily to fix the rescale parameters \( c_1 \) and \( c_2 \) for the purpose of phase classification.

![Graph](image)

**FIG. S7. Rescaling the predicted caging order parameters.** (a) Determination of \( c_1 \) and \( c_2 \) for the melting transition. The data points of \( q_i \) and \( \Delta_i \) are collected from \( N = 125 \) particles in \( N_w = 64152 \) samples in the temperature window \( T \in (0.033, 0.123) \). The line represents the best fitting, \( \Delta_i = c_1 q_i - c_2 \), where \( c_1 \) and \( c_2 \) are fitting parameters. (b) Comparison of the averaged prediction \( q = \frac{1}{2} \sum q_i \) (points) and the physical order parameter \( \Delta \) (line, see Eq. (5) for the definition), at different volume fractions \( \varphi \) around melting. We perform 12 independent runs (the initial values of network parameters are the same), and plot the obtained \( c_1 \) and \( c_2 \) in the inset.

### B. Caging order parameters learned by first-level networks

In the main text, we show that the rescaled output \( \hat{q}_i = c_1 q_i - c_2 \) from the first-level small networks agree with the physically defined caging order parameters, i.e., \( \Delta_i \) for the melting transition (see Fig. 3), or \( \chi_i \) for the Gardner transition (see Fig. 5), where \( \hat{q}_i \) is the direct output. To determine the parameters \( c_1 \) and \( c_2 \), we collect the pairs of \( q_i \) and \( \Delta_i \) (or \( \chi_i \)) of all \( N \) particles in \( N_w \) samples at \( N_f \) different temperatures (covering both phases), and perform a linear fitting (see Fig. S7(a)) for an

![Graph](image)

**FIG. S8. Independence of \( \hat{T}_m \) and \( w \) on the blanking window \([\hat{T}_2, \hat{T}_1]\), for the melting transition.** The two-level NNN is trained using a few different combinations of \( \hat{T}_1 \) and \( \hat{T}_2 \), for \( N = 2000 \) and \( \Gamma = 10^{-4} \). The predicted \( \hat{T}_m \) and \( w \) are plotted as functions of \( \hat{T}_{\text{center}} \) and \( \Delta \). The horizontal dashed lines are \( \hat{T}_m(N = 2000) = 0.064 \) and \( w(N = 2000) = 0.001 \) obtained for \( \hat{T}_1 = 0.083 \) and \( \hat{T}_2 = 0.053 \). The solid line in (a) represents \( \hat{T}_m = \hat{T}_{\text{center}} \).

### C. Independence of learning results on the blanking window

During training, the samples at \( \hat{T} > \hat{T}_1 \) and \( \hat{T} < \hat{T}_2 \) are labeled as in the liquid and glass phases respectively. The samples in the blanking window \([\hat{T}_2, \hat{T}_1]\) are not used. Figure S8 shows that the machine predicted melting temperature \( \hat{T}_m \) and transition width \( w \) are independent of the blanking window (more specifically, the center of window \( \hat{T}_{\text{center}} = (\hat{T}_1 + \hat{T}_2)/2 \) and the width of window \( \Delta \hat{T} = \hat{T}_1 - \hat{T}_2 \)). Note that, obviously we should require \( \hat{T}_m \) to be inside of the blanking window, i.e., \( \hat{T}_2 < \hat{T}_m < \hat{T}_1 \). With this restriction, the choice of blanking window is flexible. For a comparison, see Fig. 9 for the situation in the glass transition.

S3. ADDITIONAL DATA FOR THE GARDNER TRANSITION

### A. Choice of input data

According to the predictions from the mean-field glass theory [27, 52], the features of stable and marginally
FIG. S9. Failure to learn the Gardner transition using \( \{\Delta'_{AB}\} \) as input data. The test is performed for \( N = 2000 \) systems.

stable phases are encoded in \( \{\Delta_i\} \), and thus in principle one should be able to use \( \{\Delta'_{AB}\} \) as input data to train networks. However, in practice, the network fails to correctly identify both phases, when \( \{\Delta'_{AB}\} \) are used as input. We find that (see Fig. S9), as \( T \to 0 \), the predicted probability \( P \approx 0.5 \), while physically we expect \( P \approx 1 \) (nearly all samples should belong to the marginally stable phase at sufficiently low \( T \)). On the other hand, correct and robust predictions are obtained when \( \{w'_{AB}\} \) (see Eq. (7)) are used as input. Therefore, it turns out that the normalization treatment in Eq. (7) can efficiently remove redundant information and noises. Indeed, the average cage size becomes smaller with decreasing \( T \), but this effect is independent of the physics of Gardner transition. The situation is different in spin systems, where the order parameters are usually always in the same range \([0,1]\) at any temperature.

![Graphical representation of the failure to learn the Gardner transition](image)

FIG. S10. Independence of \( \hat{T}_G \) and \( w \) on the blanking window \([T_2, T_1]\), for the Gardner transition. The predicted \( \hat{T}_G \) and \( w \) are plotted as function of \( T_{\text{center}} \) and \( \Delta T \) (for \( N = 8000 \) systems). The horizontal dashed lines are \( \hat{T}_G(N=8000) = 0.0067 \) and \( w(N=8000) = 0.0012 \) obtained for \( T_1 = 0.011 \) and \( T_2 = 0.0045 \). The solid line in (a) represents \( \hat{T}_G = \hat{T}_{\text{center}} \).

![Graphical representation of the independence of learning results](image)

B. Independence of learning results on the blanking window

Figure S10 shows that the machine predicted Gardner transition temperature \( \hat{T}_G \) and the transition width \( w \) are independent of the choice of blanking window. In contrast, parameter-dependent results are found for the glass transition (see Fig. 9).

![Graphical representation of finite-size scaling](image)

FIG. S11. Finite-size scaling of machine learned transition width \( w \) of the Gardner transition. The lines represent Eq. (52), where \( b = 1/2 \) for solid blue, \( b = 1/d\nu = 0.45(1) \) (\( \nu = 0.74(2) \)) for solid red, and \( b = 1/d\nu = 0.43(1) \) (\( \nu = 0.78(2) \)) for dashed red (data from Ref. [42]).

![Graphical representation of the Gardner transition temperature](image)

FIG. S12. Comparison of the Gardner transition temperature \( \hat{T}_G \). We plot machine predicted \( \hat{T}_G(N) \) obtained in this work and from Ref. [42], as well as \( \hat{T}_G(N = 1000) \approx 0.0078 \) from a physical method [17].
C. Further comparisons

In Fig. S5, we have shown that a finite-size analysis of machine learning results, following Eq. (2) or Eq. (S2), can be used to distinguish between first- and second-order phase transitions in the Ising model. In Fig. S11, we repeat such a comparison for melting and Gardner transitions. The data of melting transition clearly follows the scaling $w \sim N^{-1/2}$ (recall that $b = 1/2$ for first-order phase transitions with disorder, see the main text), confirming its discontinuous nature. On the other hand, a deviation from this scaling, beyond numerical noises, can be seen in the data of Gardner transition. Thus the finite-size analysis suggests that the Gardner transition is continuous. This conclusion is confirmed by the unsupervised learning (Fig. 7), which identifies two states in the marginal glass phase below $T_G$. We further compare our results to those obtained by previous machine learning [42] and physical [17] methods (Figs. S11 and S12). Note that Ref. [42] used a single FNN and $\{\chi_i\}$ as input. The approach presented here is more general since it can automatically extract $\{\chi_i\}$ from nested networks.

S4. ADDITIONAL DYNAMICAL DATA FOR THE GLASS TRANSITION

In Fig. S13, we plot the data of mean-squared displacement (MSD)

$$\delta r^2(t) = \frac{1}{N} \sum_{i=1}^{N} |r_i(t) - r_i(0)|^2,$$

where $r_i(0)$ is the position of particle $i$ right after compression with a rate $\Gamma = 10^{-3}$, and $r_i(t)$ is the position at time $t$ (we set $t = 0$ and $\Gamma = 0$ after compression). The dynamics clearly slow down with decreasing $\hat{T}$, but activated processes are non-negligible since the MSD is not completely flat even at low $\hat{T}$ (for a comparison, see the MSD of ultra-stable glasses in Fig. 2 of Ref. [17]). Figure S13(b) shows that the caging order parameter $\Delta$ changes smoothly with $\varphi$ and depends sensitively on the measurement time $t$. In short, the glass transition is not a sharp phase transition from the dynamical point of view, and not inconsistently, our machine learning method fails to determine a unique transition point (Fig. 9). In contrast, although the finite-time effect exists in the Gardner transition, robust results of both the transition point and the critical scaling can be obtained through a combined finite-size-finite-time scaling analysis [42].