The Journal of Chemical Physics

Dimensionality reduction of local structure in glassy binary mixtures

Cite as: J. Chem. Phys. **157**, 204503 (2022); https://doi.org/10.1063/5.0128265
Submitted: 27 September 2022 • Accepted: 04 November 2022 • Published Online: 22 November 2022

Daniele Coslovich, Robert L. Jack and Joris Paret

View Online Export Citation CrossMark

AIP Publishing

Dimensionality reduction of local structure in glassy binary mixtures

Cite as: J. Chem. Phys. 157, 204503 (2022); doi: 10.1063/5.0128265
Submitted: 27 September 2022 • Accepted: 4 November 2022 •
Published Online: 22 November 2022

Daniele Coslovich,[1,a)] Robert L. Jack,[2,3] and Joris Paret[4]

AFFILIATIONS

[1] Dipartimento di Fisica, Università di Trieste, Strada Costiera 11, 34151 Trieste, Italy
[2] Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom
[3] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom
[4] Laboratoire Charles Coulomb, Université de Montpellier, Montpellier, France

Note: This paper is part of the JCP Special Topic on Machine Learning Hits Molecular Simulations.
[a)] **Author to whom correspondence should be addressed:** dcoslovich@units.it

ABSTRACT

We consider unsupervised learning methods for characterizing the disordered microscopic structure of supercooled liquids and glasses. Specifically, we perform dimensionality reduction of smooth structural descriptors that describe radial and bond-orientational correlations and assess the ability of the method to grasp the essential structural features of glassy binary mixtures. In several cases, a few collective variables account for the bulk of the structural fluctuations within the first coordination shell and also display a clear connection with the fluctuations of particle mobility. Fine-grained descriptors that characterize the radial dependence of bond-orientational order better capture the structural fluctuations relevant for particle mobility but are also more difficult to parameterize and to interpret. We also find that principal component analysis of bond-orientational order parameters provides identical results to neural network autoencoders while having the advantage of being easily interpretable. Overall, our results indicate that glassy binary mixtures have a broad spectrum of structural features. In the temperature range we investigate, some mixtures display well-defined locally favored structures, which are reflected in bimodal distributions of the structural variables identified by dimensionality reduction.

I. INTRODUCTION

Short-range order in liquids and glasses is evidenced by a sharp peak in the radial distribution function, which defines the first coordination shell. However, the spatial arrangement of atoms and molecules within this coordination shell is more difficult to characterize.[1,2] Conventional approaches to describe local order beyond the two-body level include higher order correlation functions,[3] bond-orientational order parameters,[4] and more general structural descriptors based on geometrical[5–7] or topological constructions.[8,9] Over the last decades, local structure analysis provided evidence that some glassy colloidal suspensions and multicomponent alloys display nontrivial local arrangements, known as locally favored structures (LFSs).[1] These structures tend to be more symmetric and stable than the bulk and also correlate to some extent with the local fluctuations of particle mobility,[10–13] i.e., dynamic heterogeneity. At present, however, there is no robust and generally accepted operational definition of a system's locally favored structure.[14]

Unsupervised learning methods have recently emerged as a promising alternative to characterize materials' local structure.[15,16] Starting from a high-dimensional descriptor of the particles' arrangements, one searches for patterns and regularities in the statistics of the descriptor. Unsupervised methods typically comprise two steps: (i) dimensionality reduction, to project the high-dimensional descriptor on a smaller subspace while retaining most of the original information, and (ii) clustering, to identify groups of points in the dataset that share similar values of the (reduced) descriptor. This approach has been recently applied, for instance, to crystal structure identification in colloidal suspensions[17,18] and the study of partially ordered systems.[16,19] Recent studies have also tackled challenging problems of structural analysis in bulk disordered materials, such as simple models of supercooled liquids and glasses,[20,21] amorphous carbon,[16,22] and liquid water in normal and supercooled conditions.[23,24] However, some of these extensions employ high-dimensional structural descriptors, which are not always easy to interpret, and involve nonlinear reduction methods, which mostly act as black boxes.

In this work, we focus on the first step of unsupervised learning, i.e., dimensionality reduction, and explore its role as a heuristic tool for structural analysis of binary glassy mixtures. Our analysis is based on smooth structural descriptors that account for both density and bond-order fluctuations. We apply principal component analysis (PCA), which is possibly the simplest linear dimensionality reduction method, to several descriptors and find evidence of a varying degree of structural heterogeneity across glassy mixtures. The first few principal components, which capture the largest variance of the original descriptor's distribution, are closely connected to physically motivated measures of local order as well as to dynamic heterogeneities. Finally, we show that a neural network autoencoder (AE), which is a more complex dimensionality reduction method, provides nearly identical results to PCA. Since linear methods have the advantage of being easily interpretable, they should be preferred when studying simple glassy systems.

The structure of this paper is as follows: Sec. II introduces our models and methods, including five different model systems, and several different descriptors that are used to characterize the local structure. Sections III–V present our key findings. Section VI gives a critical discussion and an outlook on the methodology.

II. METHODS

We consider a pool of five computational models of binary glass-forming liquids, characterized by different types and degrees of local order: two canonical computer models, the Kob-Andersen[25] (KA) mixture and the Wahnström[26] (Wahn) mixture, two Lennard-Jones models that mimic the structure of amorphous $Ni_{33}Y_{67}$[27] and SiO_2,[28] respectively, and a realistic embedded-atom model of $Cu_{64}Zr_{36}$.[29] We analyze statistically uncorrelated configurations, separated by at least one structural relaxation time τ_α, obtained from classical molecular dynamics simulations. As usual, τ_α is defined by the condition $F_s(k^*, \tau) = 1/e$, where $F_s(k^*, t)$ is the self-intermediate scattering of the system at a wave-vector k^* corresponding to the first peak of the structure factor. To provide a consistent comparison, all the systems are studied under equilibrium conditions at temperatures close to their respective mode-coupling theory (MCT) crossover temperatures, empirically identified by a power law fit of the structural relaxation time data.[30] Around the MCT crossover, the structural relaxation times are typically about 3 orders of magnitude larger than at the onset of slow dynamics. Full details about the models are given in Appendix A.

For each of these models, we analyze several structural descriptors, ranging from simple bond-order (BO) parameters and their smooth variants, to higher-dimensional descriptors that account for both orientational and radial correlations. To reduce the dimensionality of these descriptors, we use both linear (PCA) and nonlinear (AE) methods. In this section, we present a synthesis of these methods and ideas. Readers already familiar with characterizations of local structure and dimensionality reduction may want to skip directly to Sec. III.

A. Structural descriptors

1. Bond order

Bond-order parameters are standard measures of structure in the first coordination shell. Let $\boldsymbol{r}_i$ be the position of particle i and define $\boldsymbol{r}_{ij} = \boldsymbol{r}_j - \boldsymbol{r}_i$ and $r_{ij} = |\boldsymbol{r}_{ij}|$. Then, consider the weighted microscopic density around particle i,

$$\rho(\boldsymbol{r}; i) = \sum_{j=1}^{N_b(i)} w_j \delta(\boldsymbol{r} - \boldsymbol{r}_{ij}), \tag{1}$$

where w_j is a particle-dependent weight and the sum involves a set of $N_b(i)$ particles, which defines the coordination shell of interest for particle i.

We project the microscopic density on a unit-radius sphere, that is, $\hat{\rho}(\hat{\boldsymbol{r}}; i) = \sum_{j=1}^{N_b(i)} w_j \delta(\boldsymbol{r} - \hat{\boldsymbol{r}}_{ij})$, where $\hat{\boldsymbol{r}} = \boldsymbol{r}/|\boldsymbol{r}|$ and similarly $\hat{\boldsymbol{r}}_{ij} = \boldsymbol{r}_{ij}/|\boldsymbol{r}_{ij}|$. Expanding in spherical harmonics yields

$$\hat{\rho}(\hat{\boldsymbol{r}}; i) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm}(i) Y_{lm}(\hat{\boldsymbol{r}}), \tag{2}$$

with coefficients

$$c_{lm}(i) = \int d\boldsymbol{r} \rho(\boldsymbol{r}; i) Y_{lm}(\hat{\boldsymbol{r}}). \tag{3}$$

In the conventional bond-order analysis, one sets the weights to unity and considers the normalized complex coefficients,

$$\begin{aligned} q_{lm}(i) &= \frac{1}{N_b(i)} \int d\boldsymbol{r} \rho(\boldsymbol{r}; i) Y_{lm}(\hat{\boldsymbol{r}}) \\ &= \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} Y_{lm}(\hat{\boldsymbol{r}}_{ij}). \end{aligned} \tag{4}$$

The rotational invariants,

$$Q_l(i) = \left(\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |q_{lm}(i)|^2 \right)^{1/2}, \tag{5}$$

provide a detailed structural description of the local environment around particle i. By truncating the expansion to order $l_{\max}$, we obtain the simplest BO descriptor of an arbitrary particle i,

$$X^{\mathrm{BO}}(i) = (Q_0(i), \ldots, Q_{l_{\max}}(i)). \tag{6}$$

In the machine learning context, such a sequence is usually referred to as a "feature vector."

We note that the complex coefficients can be averaged over nearest neighbors, as suggested by Lechner and Dellago,[31] to provide an improved descriptor for crystal structure detection. We have tested this approach, but we do not use it in this work because we found that the additional average tends to smear the differences between disordered structural environments in the systems of our interest.

2. Neighbor definition and smoothed BO parameters

In the following, we mostly focus on the structural heterogeneity (or diversity of local arrangements) within the first coordination shell. It is therefore important to discuss how this coordination shell is defined.

The most common approach is to define the neighbors on the basis of a fixed cutoff distance $r^c_{\alpha\beta}$, where α and β are species indices. The cutoff distance is equal to the first minimum of the corresponding partial radial distribution functions $g_{\alpha\beta}(r)$. Alternative definitions include setting the neighbors on the basis of a (radical) Voronoi tessellation[6] or using the solid angle nearest neighbor (SANN) approach,[32] which is parameter-free.

Irrespective of this choice, the coefficients Q_l will change discontinuously whenever a particle leaves the coordination shell. The fact that the descriptor's components are not smooth functions of particles' coordinates is a concern.[33] The issue becomes particularly serious at low temperatures, when thermal fluctuations are small and discontinuities may affect the distribution of the BO parameters. An obvious approach to counter this issue is to use a smeared local density, in which the Dirac deltas are replaced by Gaussians,[34,35] see Sec. II A 3. Alternatively, or even in addition to Gaussian smearing, one can consider a smooth version of the descriptor, in which the coefficients q_{lm} are multiplied by a weighting function $f(r)$ that depends on the radial distance r between the central particle and its neighbors,

$$q^S_{lm}(i) = \frac{1}{Z(i)}\sum_{j=1}^{N} f(r_{ij})Y_{lm}(\hat{\boldsymbol{r}}_{ij}), \tag{7}$$

where $Z(i) = \sum_{j=1}^{N} f(r_{ij})$ is a normalization constant and the superscript S indicates the smooth nature of the descriptor. In the following, we use

$$f(r_{ij}) = \exp\left[-(r_{ij}/r^c_{\alpha\beta})^{\gamma}\right]H(R^c_{\alpha\beta} - r_{ij}), \tag{8}$$

where $r^c_{\alpha\beta}$ is the first minimum of the corresponding partial radial distribution function for the pair (i,j) and γ is an integer whose value is given in Sec. III. Moreover, H is the Heaviside step function, which ensures, for efficiency reasons, that the descriptor only has contributions from particles within a distance $R^c_{\alpha\beta} = 1.3 \times r^c_{\alpha\beta}$ from the central one.

The rotational invariants are defined similar to Eq. (5) and the corresponding "smooth" bond-order (SBO) descriptor of an arbitrary particle i is given by

$$X^{\mathrm{SBO}}(i) = (Q^S_0(i), \ldots, Q^S_{l_{\max}}(i)). \tag{9}$$

We note that in network-forming systems, such as amorphous silica, the first coordination shell is sharply defined at low temperatures. In these systems, the breakage of a bond leads to a "genuine" discontinuity in the BO parameters. This suggests to use a smoothing function that decays rapidly around $r^c_{\alpha\beta}$, i.e., a high value of γ, to avoid a too strong smearing of the local environment. Moreover, in mixtures with strong chemical order due to covalent bonding, it is appropriate to compute separate descriptors depending on the species of the neighboring particles.[16] For the silica model introduced in Appendix A 3, we therefore restrict the sum in Eq. (7) to particles whose species is distinct from the one of the central particle i.

3. Radial dependence of bond order: SOAP

By projecting the local density on a unit-radius sphere, the conventional BO order analysis treats the first coordination shell as a whole and ignores any radial dependence. Modern machine learning descriptors, used for instance to fit potential energy surfaces, provide instead a systematic expansion of the local density.[36] Here, we will focus on the smooth overlap of atomic positions (SOAP) descriptor,[34] which complements the spherical harmonics with a radial basis $g_n(r)$. One can thus analyze the BO at different length scales as well as correlations between BO orders at different distances.[34] We will use the implementation of the SOAP descriptor provided by the `DScribe` package,[37] whose documentation provides a wealth of related information.

Within the SOAP descriptor, the microscopic density is smeared with Gaussians of width σ,

$$\rho(\boldsymbol{r}; i) = \sum_{j=1}^{N_b(i)} w_j \exp\left(-\frac{|\boldsymbol{r} - \boldsymbol{r}_{ij}|^2}{2\sigma^2}\right), \tag{10}$$

and w_j is set to 1. Assuming that the radial basis functions $g_n(r)$ are orthonormal, the expansion reads

$$\rho(\boldsymbol{r}; i) = \sum_{n=1}^{n_{\max}}\sum_{l=0}^{l_{\max}}\sum_{m=-l}^{-l} c_{nlm}(i)\, g_n(r) Y_{lm}(\hat{\boldsymbol{r}}), \tag{11}$$

with

$$c_{nlm}(i) = \int d\boldsymbol{r}\rho(\boldsymbol{r}; i) g_n(r) Y_{lm}(\hat{\boldsymbol{r}}). \tag{12}$$

Notice that, contrary to the conventional BO descriptor—see Eq. (4)—the coefficients $c_{nlm}(i)$ are *not* normalized by the number of neighbors. This means that the descriptor explicitly accounts for the "coordination number" $N_b(i)$ around a particle, i.e., how dense is the local coordination shell. We will further discuss this point in Sec. III C.

The basic SOAP descriptor is then defined by the power spectrum $p_{nl}(i) = \sum_{m=-l}^{l} c^*_{nlm}(i) c_{nlm}(i)$.[34] However, this quantity is not very sensitive to angular correlations between particles at different distances, so that a faithful description of the local particle environment requires a more general descriptor that includes such correlations explicitly.[34] A suitable choice is given in Ref. 38,

$$Q_{nn'l}(i) = \left(\frac{8\pi^2}{2l+1}\right)^{1/2} \sum_{m=-l}^{l} c^*_{nlm}(i) c_{n'lm}(i). \tag{13}$$

This descriptor retains a lot more information than the single power spectrum $p_{ln}(i)$, which only includes the diagonal terms in the radial basis expansion, but is of course also computationally more expensive.

To sum up, the full SOAP descriptor of an arbitrary particle i is defined by the following feature tensor:

$$(\ldots, Q_{nn'l}(i), \ldots), \tag{14}$$

with $0 \le l \le l_{\max}$, $1 \le n \le n_{\max}$, and $n' \ge n$, i.e., dropping the terms that are identical by symmetry. In the following, we will flatten this tensor as a vector to form the descriptor $X^{\mathrm{SOAP}}(i)$ of an arbitrary particle i. We will use the original radial basis suggested in Ref. 34,

$$g_n(r) = \sum_{n'=1}^{n_{\max}} \beta_{nn'}(r - r_{\mathrm{cut}})^{n'+2}, \tag{15}$$

where the coefficients $\beta_{nn'}$ ensure orthonormality. These basis functions are not associated with neighbors at specific distances: This helps to ensure a faithful description of the local environment at the cost of reducing the interpretability of the descriptor. Note that choosing $n_{\max} = 1$ provides a smooth version of the conventional BO descriptor, which differs nonetheless from the SBO descriptor because of the Gaussian smoothing of the local density and the lack of normalization.

4. Radial dependence of bond order: A simpler Gaussian basis

It is interesting to connect the SOAP descriptor to the one used by Boattini *et al.*[39] in a recent supervised learning study of dynamic heterogeneity. We will refer to their descriptor as radial bond-order (RBO) descriptor because it captures the radial dependence of bond order in the most straightforward way. This descriptor does not involve any smoothing of the local density. As a radial basis, Boattini *et al.* used Gaussian functions of width δ centered on a grid of distances $\{d_n\}_{n=1\ldots n_{\max}}$,

$$G_n(r) = \exp\left(-\frac{(d_n - r)^2}{2\delta^2}\right). \tag{16}$$

As noted above, such descriptors are not very sensitive to angular correlations between particles at different distances.[34] It is unclear *a priori* to what extent this issue will affect supervised and unsupervised learning of structure and dynamics via this descriptor.

The complex radial bond-order coefficients are defined as

$$q^R_{lmn}(i) = \frac{1}{Z(i)} \sum_{j=1}^{N} G_n(r_{ij}) Y_{lm}(\hat{\mathbf{r}}_{ij}), \tag{17}$$

where $Z(i) = \sum_{j=1}^{N} G_n(r_{ij})$ is a normalization constant and the superscript R indicates the radial dependence of the descriptor. In the following, we actually use

$$G_n(r_{ij}) = \exp\left(-\frac{(d_n - r_{ij})^2}{2\delta^2}\right) H(R_{\max} - r_{ij}), \tag{18}$$

where H is the Heaviside step function, which allows us once again to neglect the contributions of particles further than a distance $R_{\max} = d_{n_{\max}} + 2.5\delta$ from the central particle, where $d_{n_{\max}}$ is the largest distance in the grid of points $\{d_n\}$. Then, only the diagonal coefficients of the power spectrum, i.e.,

$$Q^R_{ln}(i) = \left(\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |q^R_{lmn}(i)|^2\right)^{1/2}, \tag{19}$$

are retained to form the descriptor of particle i as $(\ldots, Q^R_{ln}(i), \ldots)$, which is again flattened as a vector $X^{\mathrm{RBO}}(i)$, composed of $l_{\max} \times n_{\max}$ structural features. The choice for the grid of points $\{d_n\}$ is discussed in Secs. III D and IV C.

One key advantage of the RBO descriptor is that it is easily interpretable: It describes the bond order of thin spherical shells at increasing distance from the central particle. The radial basis considered within SOAP are much less interpretable and the descriptor involves the full power spectrum, $p_{nn'l}$. This provides a wealth of information, but it is unclear *a priori* whether this is actually relevant for a specific problem. We will come back to this issue in Sec. IV A.

Finally, we note that the RBO descriptor does not explicitly account for the local coordination number (i.e., how dense is the shell of neighbors) because the projected density does not scale with the number of neighbors. To include this information, Boattini *et al.* have complemented their structural descriptor with a measure of the local density in successive shells centered around a given particle. Interestingly, however, we will find that the normalized BO descriptors are correlated with the local density of the first coordination shell (Sec. III C).

B. Dimensionality reduction

We have described the construction of several structural descriptors $X(i)$ that characterize the first coordination shell of particle i as given in Eqs. (6), (9), and (19). Let the dimension of $X(i)$ be M. For a sample of $N_{\mathrm{tot}} = N \times n_{\mathrm{conf}}$ particles, this means that the coordination shells of all particles can be summarized in a matrix of size $N_{\mathrm{tot}} \times M$, whose rows are the $X(i)$,

$$\mathbf{X} = \begin{pmatrix} X(1) \\ \vdots \\ X(N_{\mathrm{tot}}) \end{pmatrix}. \tag{20}$$

In a liquid, each row of the matrix is different, which reflects the heterogeneity of the local structure. However, correlations within the coordination shell mean that the row vectors have a nontrivial probability distribution, with correlations among their components. The goals of dimensionality reduction are (i) to find out in a generic way to what extent the components are correlated, (ii) to extract a reduced descriptor of lower dimension, which embodies most of the "relevant" fluctuations, and (iii) to provide quick insight into structural heterogeneity, such as multimodality and non-Gaussianity.

In simple systems and with simple descriptors, such as X^{BO}, one may always reduce the dimensionality of X by focusing on selected components, e.g., Q_4 and Q_6, perhaps motivated by some physical intuition about the local structure. One may then use the low-dimensional distributions of the chosen components to characterize structural heterogeneity. The aim of unsupervised learning is to avoid assumptions on the most important components[17,18] and to identify them using statistical methods. This is particularly advantageous when M is large and even more so when applying unsupervised learning methods that require computing distances between datapoints.[40]

Dimensionality reduction is a valuable tool to gather the relevant components into "reduced descriptors." It involves mapping the high-dimensional descriptors into a lower-dimensional space,

$$\mathcal{V}_{\mathbf{X}} : X \mapsto \widetilde{X}, \tag{21}$$

where $\mathcal{V}_{\mathbf{X}}$ is a function that maps the original M-dimensional descriptor [e.g., $X(i)$] to a reduced P-dimensional descriptor [here $\widetilde{X}(i)$], with $P < M$. The method exploits the full matrix $\mathbf{X}$ to construct the mapping, typically by optimizing some cost function; hence, $\mathcal{V}_{\mathbf{X}}$ carries an implicit dependence on $\mathbf{X}$. The main difference between the various dimensionality reduction methods lies in the nature of the mapping: We talk about *linear* or *nonlinear* dimensionality reduction depending on the linearity of $\mathcal{V}_{\mathbf{X}}$, as will be discussed further in the next paragraphs.

Note that the dataset is often normalized before performing dimensionality reduction.[15] This normalization, known as feature scaling, is necessary if the dataset is heterogeneous, with different features having different physical dimensions. In this work, the features have the same dimensions, so scaling is not strictly necessary. In some cases, we perform a simple Z-score scaling: Each feature is normalized by subtracting the average and dividing by the standard deviation evaluated over all the particles. We indicate whether or not feature scaling is applied for each descriptor.

1. Principal component analysis

Principal component analysis (PCA)[41] is a linear dimensionality reduction method that aims to transform each descriptor X into $\widetilde{X}$ by retaining directions associated with the largest variance of the dataset. Since the transformation is linear, this is easily performed at the level of the data matrix $\mathbf{X}$, which is first transformed as

$$\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{V}, \tag{22}$$

where $\mathbf{V}$ is a $M \times M$ matrix whose columns $\{V^{(j)}\}_{j=1\ldots M}$ are called the principal component (PC) directions of $\mathbf{X}$.[42] They correspond to the eigenvectors of the features' covariance matrix, whose entry (j,k) is

$$\mathrm{cov}(X_j, X_k) = \langle (X_j - \langle X_j \rangle)(X_k - \langle X_k \rangle) \rangle, \tag{23}$$

where $\langle \ldots \rangle$ denotes an average over the N_{tot} particles of the sample. This covariance matrix measures correlations between different elements of the descriptor. The eigenvectors $\{V^{(j)}\}_{j=1\ldots M}$ correspond to the directions of the new feature space. Their associated eigenvalues $\{\lambda_j\}_{j=1\ldots M}$ indicate the variance of the data along the corresponding eigenvector, and the explained variance ratio (EVR) of the j-th principal component is thus given by

$$\mathrm{EVR}(\mathrm{PC}_j) = \frac{\lambda_j}{\sum_{k=1}^{M} \lambda_k}. \tag{24}$$

By convention, the principal components are sorted in descending order according to their EVR.

Dimensionality reduction is performed by considering a truncated transformation, which retains only the first $P < M$ eigenvectors in the matrix $\mathbf{V}$. Equivalently, we retain only the first P columns of the transformed matrix $\widetilde{\mathbf{X}}$, yielding P reduced features, $\widetilde{X}_1, \ldots, \widetilde{X}_P$. As a rule of thumb, one keeps as many PC directions as needed to explain to a significant fraction, say 80%, of the total variance of the dataset, see also Sec. III C. Alternatively, one can inspect the EVR as a function of the component index and look for an inflection point, which defines empirically an optimal P for dimensionality reduction. The notion of intrinsic dimension[43,44] of a dataset provides a more robust and objective criterion. This kind of analysis is, however, beyond the scope of the present work and is left for a future study. Note that within PCA, the reduced features are linearly related to the original ones. Thus, one can directly identify the main sources of variance in the data, e.g., which features are most responsible for the observed structural heterogeneity of the sample.

2. Neural network autoencoder

An autoencoder is an artificial neural network that learns an efficient coding of a dataset $\mathbf{X}$ in an unsupervised way.[42] It is trained to reproduce its own M-dimensional input as output by forcing the input data through a lower P-dimensional bottleneck in the hidden layers. An autoencoder is separated into two parts: (i) an *encoder* (the first half of the hidden layers up to the bottleneck) that forces a compression of the input data and (ii) a *decoder* (the second half of the hidden layers), tasked to reconstruct the input $\mathbf{X}$ as output $\mathbf{X}^R$ with the highest accuracy. This reduction method is nonlinear since, in feed-forward neural networks, nonlinear functions are generally used to control neuron activations between successive hidden layers.

Once the network has been optimized to minimize the reconstruction error, cutting it at the level of the encoder yields a mapping $\mathcal{V}_{\mathbf{X}}$ from the original high-dimensional data to a low-dimensional representation $\widetilde{\mathbf{X}}$. The nonlinearity of the method makes it possible to provide low-dimensional representations of datasets with intricate structures. However, an autoencoder also requires tuning a large number of parameters, such as the dimensions of the network (i.e., number and sizes of the hidden layers), choosing the activation functions, the cost function and regularization term, the solver, etc. Moreover, training the neural network is a stochastic procedure that may yield very different outcomes and must be repeated until the optimal reconstruction error on $\mathbf{X}$ is reached.

The EVR can be determined by computing the mean squared error on the reconstruction rescaled by the mean squared deviation of the input vectors,[20]

$$\mathrm{EVR} = 1 - \frac{\sum_{i=1}^{N_{\mathrm{tot}}} \|X(i) - X^R(i)\|^2}{\sum_{i=1}^{N_{\mathrm{tot}}} \|X(i) - \langle X \rangle\|^2}, \tag{25}$$

where $X^R(i)$ is the i-th row of the reconstructed input and $\langle X \rangle$ is the mean input vector. Note that this EVR is the fraction of the variance that is captured by the reduced dataset $\widetilde{\mathbf{X}}$ of the AE, while the corresponding quantity, Eq. (24), for PCA is the variance captured along a single PC direction.

III. DIMENSIONALITY REDUCTION OF LOCAL STRUCTURE

In this section, we discuss some of the key features that emerge from dimensionality reduction of bond order in glassy binary mixtures. The computation of descriptors and most of the dimensionality reduction analysis reported in the following sections were performed using the `partycls` package.[45] Data analysis has been carried out using a reproducible workflow, deposited in the Zenodo public repository.[46]

We focus here on the local structure around the small particles of the "close-packed" mixtures (Wahn, KA, $\mathrm{Ni}_{33}\mathrm{Y}_{67}$, $\mathrm{Cu}_{64}\mathrm{Zr}_{36}$) and on the Si particles of the SiO_2 model. Local order around particles of these species becomes well-defined at low temperature. We present some results for the other species in Appendix B. We mostly limit ourselves to the first coordination shell and we set the weights w_j to unity in Eq. (1). We work at temperatures that correspond to moderate supercooling, i.e., close to the MCT crossover temperature, and we analyze instantaneous configurations from dynamical simulations. (At these temperatures, analysis of energy-minimized and instantaneous configurations differ only marginally.)

A. Illustrative example: Icosahedral local order in the Wahn model

To illustrate the key features of the PCA approach to dimensionality reduction, we start by analyzing the Wahn mixture. This model serves as a useful benchmark because its local structure has been characterized in depth and is known to display strong icosahedral order.[10,11] Its structural features should thus stand out most clearly.

We consider the SBO descriptor, which avoids the discontinuous dependence of the features on the particle coordinates, taking $l_{\max} = 8$ and $\gamma = 8$ for the smoothing function $f(r)$ in Eq. (8). Using PCA for dimensional reduction, we write $\widetilde{X}_j(i)$ as the j-th element of $\widetilde{X}(i)$, which represents the projection of particle i's descriptor on the j-th PC direction. By considering all the particles in the system, we form a joint probability distribution of the PCs $(\widetilde{X}_1, \widetilde{X}_2, \ldots)$. Figure 1(a) shows the probability density function $p(\widetilde{X}_1, \widetilde{X}_2)$ of the projections on the two first PCs' directions. Since the values of Q_l^S (and Q_l) are directly comparable in terms of both amplitudes and ranges, we do not perform feature scaling.

By visual inspection, we recognize two modes in the distribution: a central diffuse region and a more localized lobe for larger values of $\widetilde{X}_1$. To illustrate the physical meaning of these modes, we compare in Fig. 1(b) the marginal distributions $p(\widetilde{X}_1)$ and $p(\widetilde{X}_1|\mathrm{LFS})$, the latter being restricted to particles at the center of the LFS as identified from a radical Voronoi tessellation.[21] It is clear that the secondary lobe is due to local icosahedral order. These observations suggest the presence of two distinct amorphous structural "states" in this model, similar to what found in polyamorphic materials.[47]

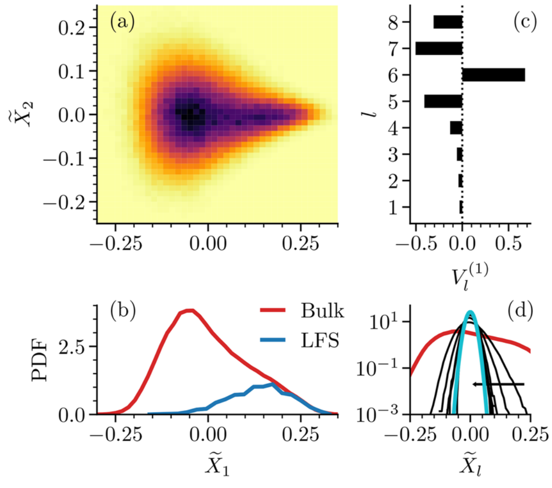

FIG. 1. Overview of the results of PCA of the SBO descriptor for the small particles in the Wahn mixture. (a) Probability density $p(\widetilde{X}_1, \widetilde{X}_2)$. (b) Marginal distribution $p(\widetilde{X}_1)$ and $p(\widetilde{X}_1|\mathrm{LFS})$. (c) Components of the eigenvector $V^{(1)}$. (d) Distribution $p(\widetilde{X}_j)$ for $j > 1$ (black) and $p(\widetilde{X}_1)$ (red). The light blue curve is a Gaussian with variance equal to the one of $\widetilde{X}_8$. The horizontal line indicates increasing PC index.

Analyzing the results in more detail, the first PC identifies the direction in feature space having the largest structural heterogeneity of the SBO parameters and accounts for about 62% of the total variance. One may expect that this component gives a strong weight to Q_6. This is confirmed by Fig. 1(c), which shows the individual components of the first eigenvector $V^{(1)}$ of the covariance matrix, see Sec. II B 1. We also see that $V^{(1)}$ couples several bond-order invariants and has large contributions also from Q_5, Q_7, Q_8, with a sign opposite to Q_6. In other words, particles with large Q_6 (and icosahedral local environments) tend to have small projections on spherical harmonics of order 5,7,8, while lower order invariants are statistically uncorrelated to Q_6.

Turning to higher principal components, we found that PC$_2$ gives strong and opposite weights to Q_5 and Q_7 (not shown) but the corresponding EVR is already fairly small (18%). Interestingly, principal components with even smaller EVR are characterized by marginal distributions $p(X_j)$ that are only slightly asymmetric and are rather well described by Gaussians, see Fig. 1(d). These observations suggest that it may be possible to schematically represent the feature space of the SBO descriptor of the Wahn mixture as having a single "relevant" direction, which captures the bulk of the structural heterogeneity, on top of a background of "trivial," nearly Gaussian directions. Interestingly, this kind of simplified picture of high-dimensional datasets defines one of the simplest models of unsupervised statistical learning[48] and is also amenable to analytical treatments.

B. From weak to strong structural heterogeneity in glassy binary mixtures

We now proceed with a more systematic comparison of the reduced structural features of glassy binary mixtures. When analyzing bond order, the vast majority of previous studies have focused on rotational invariants of order 4 and 6, which are assumed to reflect the relevant symmetries of close-packed local structures.[49] The previous example shows that PCA identifies indeed Q_6 as the most relevant parameter in the Wahn mixture. Here, however, we also consider systems for which the relevant symmetries are not obvious from physical intuition, so that unsupervised learning of local structure is useful.

As in Sec. III A, we use the SBO descriptor with $l_{\max} = 8$ and $\gamma = 8$ for the smoothing function. We checked that inclusion of higher order spherical harmonics ($l_{\max} = 16$) does not appreciably change the results. We focus on the two PC directions with the largest eigenvalues, further analysis of the remaining components is given in Sec. III C below. In Fig. 2, we show the probability density functions $p(\widetilde{X}_1, \widetilde{X}_2)$ for all the studied systems. We start by analyzing the KA mixture and the $\mathrm{Ni}_{33}\mathrm{Y}_{67}$ model. In these systems, the first two PCs capture 37% + 31% (KA) and 39% + 29% ($\mathrm{Ni}_{33}\mathrm{Y}_{67}$) of the total variance. The eigenvectors $V^{(1)}$ and $V^{(2)}$ of each model are given in Appendix B. Even though the invariants Q_4 and Q_6 feature prominently in the first two PC directions, the fluctuations of other invariants are more relevant in some models. For instance, in the KA mixture and $\mathrm{Ni}_{33}\mathrm{Y}_{67}$ model, Q_5 has the largest contribution to PC$_1$. As can be seen in Figs. 2(a) and 2(b), these two models display a broad but unimodal distribution of the reduced descriptor $(\widetilde{X}_1, \widetilde{X}_2)$. These results can be contrasted to what we found in the Wahn mixture and the $\mathrm{Cu}_{64}\mathrm{Zr}_{36}$ model, which display instead a nearly bimodal $p(\widetilde{X}_1, \widetilde{X}_2)$.

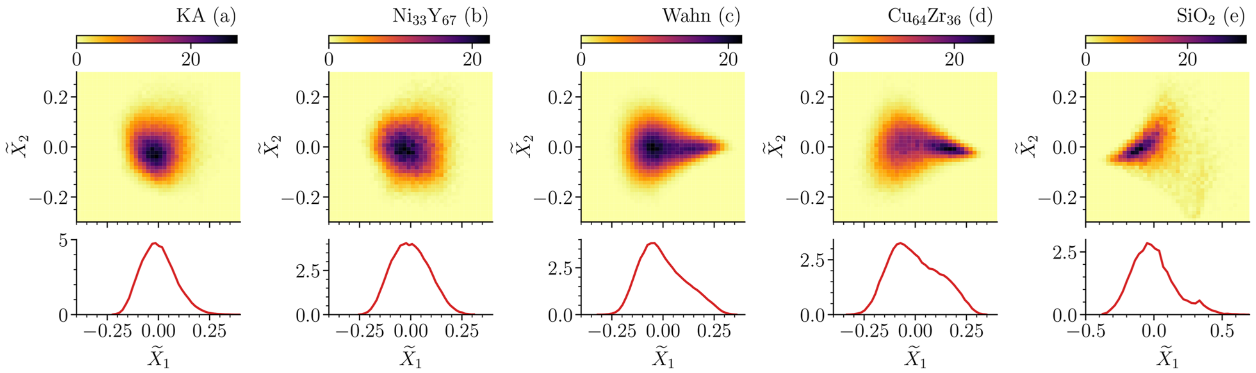

FIG. 2. Probability density $p(\widetilde{X}_1, \widetilde{X}_2)$ from the PCA of the SBO descriptor in all the studied models. Results are shown for the small particles in panels (a)–(d) and for Si particles in (e). The bottom panels show the marginal distributions of $\widetilde{X}_1$.

It is interesting to compare these results to those obtained without smoothing the descriptor, i.e., with the bare BO descriptor of Eq. (6), see Fig. 3. This descriptor has been used in a recent unsupervised learning analysis of glassy systems.[20] The PCA of X^{BO} reveals clear banding in the probability density of the reduced descriptor of all the studied models. As shown in Fig. 4, these bands are associated with different coordination numbers and disappear at high temperature, when thermal noise is large enough to smear the discontinuities of the descriptors.[50] Note that small signs of discontinuities are visible at the level of the distributions of the individual invariants, but they become much more visible after dimensionality reduction.

The fact that the reduced BO descriptor is strongly modulated by coordination number may impact the cluster analysis of these structural descriptors. However, we think these findings differ from "structure from chance" artifacts that may affect unsupervised

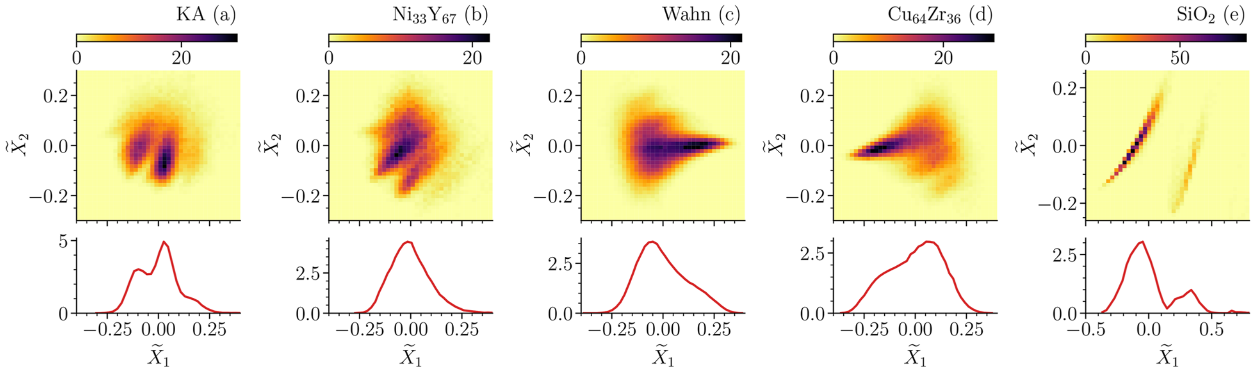

FIG. 3. Same as Fig. 2 but panels (a)–(e) correspond to the BO descriptor.

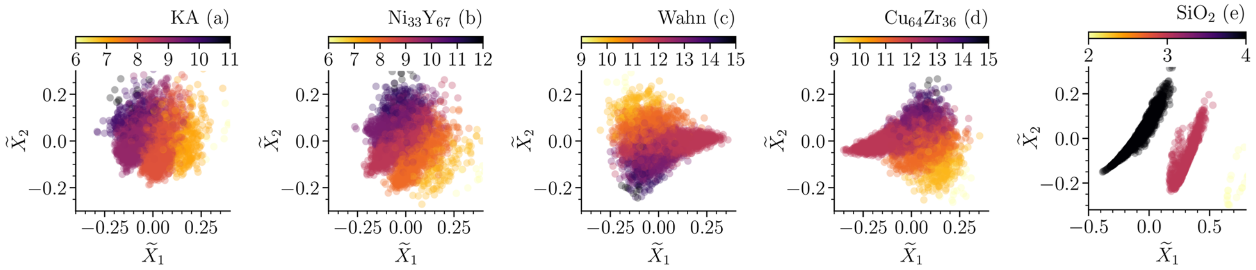

FIG. 4. Scatter plot of $(\widetilde{X}_1, \widetilde{X}_2)$ obtained from the PCA of the BO descriptor, color-coded by coordination number z. Results are shown for the small particles in panels (a)–(d) and for Si particles in (e). For clarity, only 3000 datapoints are shown for each system.

learning.[48] The fact that bond order depends on the coordination number (or more generally local density) makes physical sense. Interestingly, as we shall see in Sec. III C, a strong correlation between the first few PCs and coordination number persists even after smoothing the descriptor.

Finally, we analyze the simple model of amorphous silica. The local structure of this system can be easily characterized by the coordination number z, which equals 4 for ideal tetrahedral structures around Si particles. At the studied density and temperature, defects appear mostly in the form of undercoordinated structures, i.e., Si particles with $z = 3$ and 2. We note that such defects would be almost completely removed by energy minimization at the studied temperature. In Figs. 2(e) and 3(e), we show the distributions $p(\widetilde{X}_1, \widetilde{X}_2)$ obtained by applying PCA to X^{SBO} and X^{BO} of Si particles, respectively. As anticipated in Sec. II A 2, we use a sharper $\gamma = 18$ exponent to smear the SBO descriptor for this model because the first coordination shell is more sharply defined than in the other mixtures, and we also restrict the calculation to oxygen neighbors.

As expected from the low coordination of the local structure, we find that the first two PC directions give strong weights to invariants with small l. The PCs obtained from the bare BO descriptor display multiple sharp bands along which $\widetilde{X}_1$ and $\widetilde{X}_2$ are strongly correlated. The bands are associated with distinct coordination numbers, i.e., $z = 4, 3, 2$ from left to right. We found that, along a given band, the local environments around Si particles are also characterized by a marked gradient of tetrahedrality, i.e., how ideal is the local tetrahedral environment, see Sec. III C. On the other hand, the projections on the PCs obtained from $\widetilde{X}^{\mathrm{SBO}}$ have a much more diffuse character. The scattered region at large $\widetilde{X}_1$ values is associated with defects in the network structure. Thus, the relevant information about the local structure of this system is somehow washed out when using the smoothed descriptor.

C. Connection of reduced structural features to other structural measures

The results discussed in Sec. III B provide some evidence of a varying degree of heterogeneity of the local structure of the studied systems. The distribution $p(\widetilde{X}_1, \widetilde{X}_2)$ ranges from broad but unimodal (KA, $\mathrm{Ni}_{33}\mathrm{Y}_{67}$) to nearly bimodal (Wahn, $\mathrm{Cu}_{64}\mathrm{Zr}_{36}$) or even possibly multimodal (SiO_2). In this section, we analyze the PCs of the SBO descriptor in detail and identify its connection with known measures of local order. This analysis will corroborate the idea that the first few PCs are indeed sufficient to grasp relevant information about local order.

First, we analyze the EVR of all the PC directions in the top panel of Fig. 5. All systems display a drop in the EVR after a few PCs: This occurs already after PC_1 in the Wahn, $\mathrm{Cu}_{64}\mathrm{Zr}_{36}$ and SiO_2 models, while a little gap is visible after PC_2 in the more weakly heterogeneous systems. This supports the view that the first two PCs alone are providing significant information about the underlying feature space. Close inspection of Fig. 5(e) suggests the existence of an additional gap after PC_3 in the amorphous silica model, see also below.

We also confirmed that none of the low-EVR variables have strongly bimodal distributions. We did so by employing Hartigans' dip test,[51] which provides a rather stringent criterion for the multimodality of a distribution. The output of the dip test is a value between 0 (unimodal distribution) and 1 (multimodal distribution). We found that the marginal distributions $p(\widetilde{X}_1)$ of the SBO descriptor in the Wahn mixture and in the $\mathrm{Cu}_{64}\mathrm{Zr}_{36}$ model have a significantly bimodal character, with dip test values close to 1. For all the other marginal distributions, the dip test value indicates a unimodal character.

We now connect the reduced structural variables $\widetilde{X}_j$ to selected measures of local order. That is, we compute the Pearson correlation coefficient R between the PCs and (i) the coordination number z, measured by integrating the partial correlation function $g_\alpha(r)$, where α is the species of interest, up to its first minimum; (ii) the Θ parameter introduced by Tong and Tanaka,[52] which measures the compactness of local environment; as in the original implementation, we used radical Voronoi neighbors for this calculation[53] and we employed the positions of the first peak of the $g_{\alpha\alpha}(r)$ radial distribution functions as measures of the effective particle diameters; for the SiO_2 model, we use instead a simple measure of tetrahedrality, i.e., Θ is defined as the average deviation of the bond angles from the ideal tetrahedral angle 109.5°; (iii) the LFS index ℓ, which equals 1 if the Voronoi polyhedron surrounding the particle corresponds to the system's LFS and 0 otherwise; the LFS has been identified from a radical Voronoi tessellation in previous work;[10,13] in the SiO_2 model, ℓ is defined as 1 if the coordination number is $z = 4$ for Si particles and 0 otherwise; (iv) local potential energy u.

The bottom panels of Fig. 5 provide an overview of these correlation coefficients for all the studied mixtures. Interestingly, the first few PCs are quite strongly connected to the local order parameters (Θ, ℓ), local potential energy (u), or coordination number (z). The degree of correlation depends on the PC and on whether the mixture is weakly or strongly heterogeneous. In particular, we notice that the correlation between PC_1 and the LFS determined from the Voronoi tessellation is significant in systems with marked structural heterogeneity (Wahn, $\mathrm{Cu}_{64}\mathrm{Zr}_{36}$, SiO_2) and less so in the other models (KA, $\mathrm{Ni}_{33}\mathrm{Y}_{67}$). Even in systems with weak structural heterogeneity, however, the projections on the first two PCs are correlated with measures of local structure, despite the more uniform spectrum of features. Interestingly, PC_2 and PC_3 in the SiO_2 model display a higher correlation than the remaining low-EVR components, confirming the above observations that these variables are structurally relevant.

We also found that these conventional structural measures are all somewhat correlated with one another (with cross-correlations in the range 0.2–0.6), capturing slightly different aspects of structural heterogeneity. The PCA of X^{SBO} allows one to capture all these features through a few collective variables, defined on the basis of structural heterogeneity alone. This suggests that the intrinsic dimension of the SBO feature space is indeed low—a more precise determination would require additional work.[44] A somewhat striking feature is that, even though the bond-order coefficients do not scale with the number of neighbors [see Eq. (4)], the first PCs are still strongly correlated with the local coordination number, z. Thus, smoothing the coefficients removes the discontinuities in the standard BO descriptor, but it does not wash out the correlation with local density evidenced in Fig. 3.

To further illustrate the advantage of considering a few PCs instead of selected SBO parameters $\{Q_l^S\}$, we computed the same structural indicators as in Fig. 5 for the bare SBO descriptor. These results are shown in Fig. 6, along with the standard deviation of each

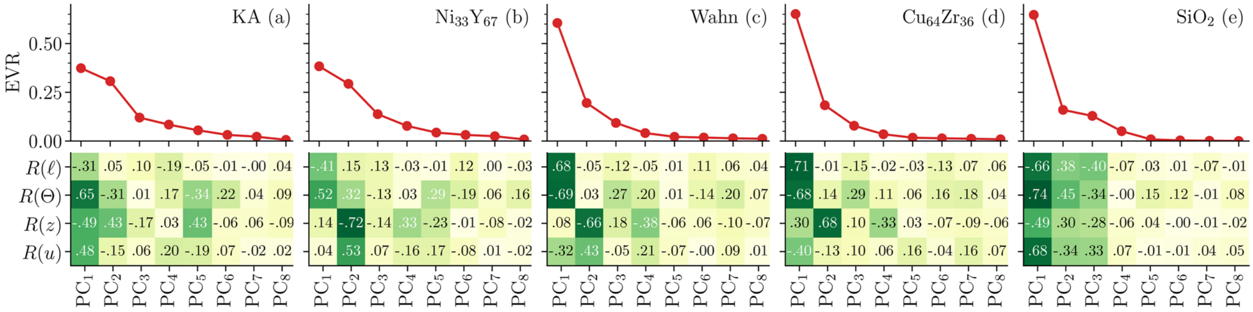

FIG. 5. Properties of the PC of the SBO descriptor for small particles in (a)–(d) and Si particles in (e). The top figures show the explained variance ratio (EVR) of each PC and the tables show the Pearson correlation coefficient R between each $\widetilde{X}_j$ and other structural measures: LFS index (ℓ), local compactness (Θ), coordination number (z) and local potential energy (u). See the main text for a detailed description of these quantities.

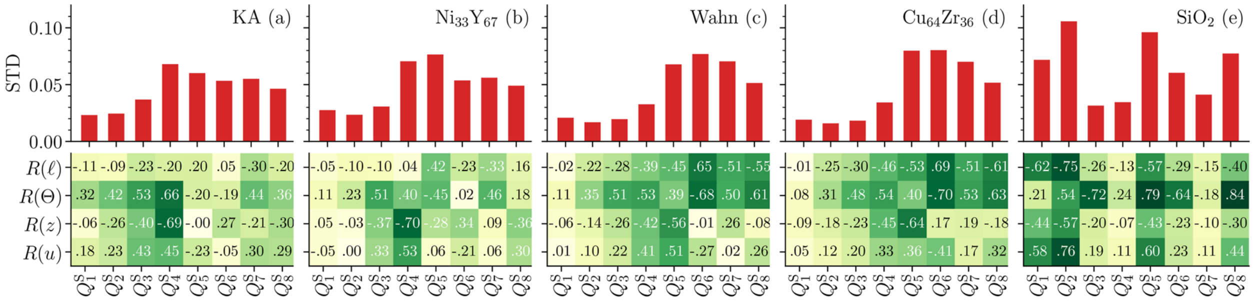

FIG. 6. Properties of the individual SBO parameters, Q_l^S, for small particles in (a)–(d) and Si particles in (e). The top panels show the standard deviation on the distributions of each parameter Q_l^S and the tables show the Pearson correlation coefficients R between the individual values of Q_l and the same structural measures featured in Fig. 5.

marginal distribution $p(Q_l^S)$. They demonstrate that, for any given system, it is difficult to pinpoint a small subset of SBO parameters that capture the bulk of structural heterogeneity. Instead, all the SBO parameters are to some extent correlated or anticorrelated to measures of local order and to local density. We note that a trivial source of correlation between z and X^{BO} (or X^{SBO}) arises from the reduction (in magnitude) of Q_l as the number of neighboring particles increases. This leads to negative correlations between most of the Q_l and z. However, in mixtures with strong icosahedral local order, Q_6 has a non-monotonic dependence with z: It peaks at $z = 12$ before dropping off at larger z. Thus, Q_6 and Q_l^S display a very weak correlation with z. Finally, we note that in the SiO$_2$ model, the first two PCs are both strongly and positively correlated with Θ. Indeed, we found a marked gradient of Θ along each of the bands visible in Fig. 3(e) (not shown).

To summarize, PCA applied to the SBO descriptor provides a few collective structural variables that aggregate information on local order. These variables are simple linear combinations of smooth BO parameters and capture the largest fraction of structural fluctuations. Our analysis has been restricted to the small particles of the close-packed mixtures, building on the insight that local order is less pronounced around the big particles (see Figs. 15 and 16 in Appendix B for analogous results for these particles). For the small particles, the first few PCs are strongly connected to well-known measures of local order and gather a signal that would otherwise be scattered across almost all the individual BO parameters. Even though some low-variance components may retain relevant information (see also Sec. IV A), our results suggest that the first PCs suffice to characterize structural heterogeneity in glassy binary mixtures, within the first coordination shell.

D. Fine-grained local structure: Using the SOAP descriptor to analyze radial dependence of bond order

Can we gain additional insight into the local structure by considering a finer-grained expansion of the local density? The SOAP descriptor provides a natural framework for this extended analysis since it performs a systematic expansion of the local density up to a distance r_{cut} from a central particle. The cost is a reduced interpretability of the descriptor and a larger hyper-parameter space to explore. Indeed, structural analysis using the SOAP descriptor must be combined with chemical and physical intuition to grasp the key results.[24] In this section, we summarize the few robust trends we have identified as well as the shortcomings of this kind of analysis.

The SOAP descriptor depends on two hyper-parameters: the smearing parameter σ in Eq. (10), which sets the length scale on which the particles are localized, and the cutoff distance r_{cut}, used to define the coordination shells of interest. In addition, it depends on the number of radial and angular basis functions $n_{\max}$

and $l_{\max}$, which control the radial and angular resolution, respectively, and on feature scaling. Unsurprisingly, we found that all these hyper-parameters have a significant impact on the outcome of dimensionality reduction, see also Ref. 16 for a discussion on SOAP hyper-parameter tuning. In this section, we again limit ourselves to the first coordination shell and we set $r^{c}_{\alpha\beta} = r_{\text{cut}} = 1.6$, slightly larger than the minimum of the total $g(r)$. This effectively ignores slight differences between coordination shells defined by particles of different species. As in the previous sections, we used $l_{\max} = 8$ orientational basis functions and varied the largest number of radial basis functions from $n_{\max} = 1$ to 8. The qualitative features discussed below are rather insensitive to this parameter for large enough $n_{\max}$. We use $n_{\max} = 6$ in the plots that follow. As a guideline, σ should be around $\sigma \approx r_{\text{cut}}/n_{\max}$ and, in thermal systems, larger than the typical vibrational amplitude of the particles. We chose $\sigma = 0.2$. Note that perturbations around $\sigma = 0.2$ may change the outcomes of dimensionality reduction, we quantify these effects later in this section.

We consider the KA and Wahn mixtures, which are representative of close-packed mixtures with weak and strong structural heterogeneity, respectively. For this analysis, we focus on a smaller dataset composed of ten configurations for each model system. In Fig. 7, we show the probability density $p(\widetilde{X}_1, \widetilde{X}_2)$ of the first two PCs of the SOAP descriptor for these two systems. In both cases, feature scaling is applied to adsorb the lack of normalization of the power spectrum.[54] The qualitative features observed in Fig. 7 are similar to those discussed in Sec. III B: $p(\widetilde{X}_1, \widetilde{X}_2)$ is nearly bimodal in the Wahn mixture, while it is unimodal in the KA mixture. Compared to the SBO descriptor, however, the EVR is now distributed more evenly among the PC directions and the first two capture a much smaller fraction of the total variance, see Fig. 8. This may be expected in view of the larger feature space of SOAP.[55] However, we can still appreciate from Fig. 8 the presence of gaps between the EVRs of the first two PCs and the bulk in the Wahn mixture: Beyond PC_2, the shape of the EVR as a function of component index evolves smoothly. In the KA mixture, a similar change is observed after 4–5 PCs, which still suggest a relatively low intrinsic dimension.[44]

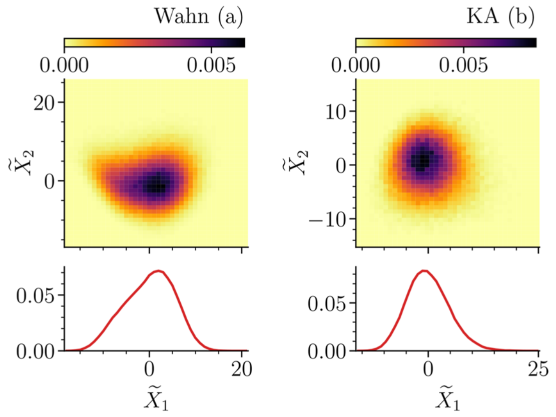

FIG. 7. Probability density $p(\widetilde{X}_1, \widetilde{X}_2)$ from the PCA of the SOAP descriptor for the small particles of (a) the Wahn mixture and (b) the KA mixture. The bottom panels show the marginal distribution $p(\widetilde{X}_1)$.

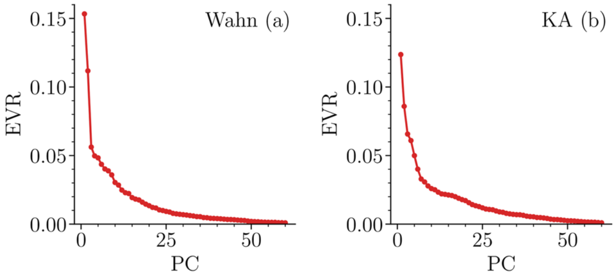

FIG. 8. EVR along the first 60 PC directions of the SOAP descriptor for the small particles of (a) the Wahn mixture and (b) the KA mixture.

However, compared to the low-dimensional SBO descriptor, these gaps are small enough to be perturbed by noise and small changes in the descriptor's hyper-parameters.

To address this point more systematically, we assessed the similarity between the reductions obtained using different choices of hyper-parameters, σ and r_{cut}. We computed the Pearson correlation coefficient R between the variable $\widetilde{X}_1$ obtained using the reference values, $\sigma = 0.2$, $r_{\text{cut}} = 1.6$, and the ones obtained for small perturbations of these hyper-parameters. From the results collected in Fig. 9, we observe that the results are relatively stable for the Wahn mixture, as expected, but the outcome of the reduction deviates significantly from the reference when r_{cut} is increased. Even more dramatic are the changes observed for the KA mixture: In this case, $\widetilde{X}_1$ may change significantly and erratically as a function of σ and r_{cut}. We attribute part of this effect to the dependence of nearest neighbor distances on the chemical composition of the first coordination shell of the KA model. It is also a general feature that smaller gaps in EVR between PCs lead to increased sensitivity of PC directions to small perturbations in the parameters or the data. This is consistent with the trend between Wahn and KA mixtures. The second PC displays an even stronger variability for both models, as expected (not shown).

Note that across the range of hyper-parameters spanned in Fig. 9, the distributions $p(\widetilde{X}_1, \widetilde{X}_2)$ maintain the same qualitative differences between the two models observed in Fig. 7. However,

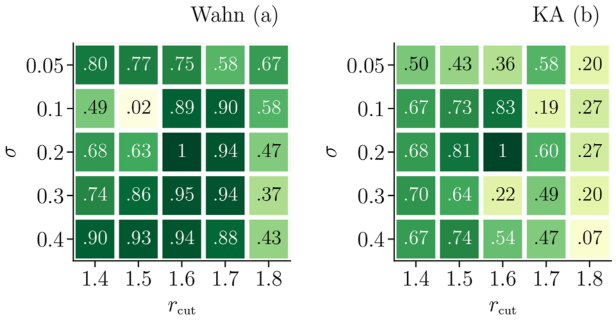

FIG. 9. Pearson correlation coefficient between the feature $\widetilde{X}_1$ of the SOAP descriptor (obtained for $\sigma = 0.2$, $r_{\text{cut}} = 1.6$) and $\widetilde{X}_1(\sigma, r_{\text{cut}})$ (obtained for different values of the parameters σ and r_{cut}, as indicated in the figure). Results are shown for the small particles of (a) the Wahn mixture and (b) the KA mixture.

specific choices of the parameters may enhance or suppress the bimodal character of $p(\widetilde{X}_1, \widetilde{X}_2)$ for the Wahn mixture. Given the large hyper-parameter space associated with the descriptor, our results call for a principled approach to hyper-parameter tuning for unsupervised learning of glassy materials, where the relevant length scales cannot be determined in a simple way.[16]

We note one possible shortcoming of the radial basis for the present analysis: As is clear from Fig. 3 of Ref. 34, the first radial functions are nonzero over a range of distances that are depleted in our range of temperatures and they vary only mildly over the first coordination shell.[56] This basis may thus require a large value of $n_{\max}$ to properly account for correlations within the first coordination shell. To test whether this could be an issue, we also used a different basis formed by Gaussian-type orbitals,[57] which effectively suppress contributions from close-contact distances. We found qualitatively similar results to those already discussed in this section. An alternative route to provide a fine-grained description of correlations within the first shell is to use the RBO descriptor, Eq. (19), which we tailored to cover the first coordination shell using $n_{\max} = 5$ Gaussians of width $\delta = 0.2$ centered at $(0.9, 1.0, 1.1, 1.2, 1.3)$, and $l_{\max} = 8$. This choice of hyper-parameters leads to PCA maps that are again qualitatively similar to those of Figs. 2(a), 2(c), and 7, but the underlying PC directions are again sensitive to addition or removal of one Gaussian (not shown here).

To sum up, SOAP has proven to be a good descriptor for supervised and unsupervised learning studies of materials and molecules,[16] in particular for fitting macroscopic properties and potential energy surfaces.[58] As we shall see in Sec. IV C, the SOAP descriptor also provides a good fit to dynamic heterogeneity, from purely structural information. Its application in the context of unsupervised learning of disordered materials is more recent and focused so far on systems with highly directional correlations, such as amorphous carbon,[16,22] or normal and supercooled liquid water.[23,24] Even in such systems with low coordination numbers and sharp orientational correlations, interpreting the results requires some physical and chemical intuition. In our opinion, the utility of this descriptor for unsupervised learning of the structure of glassy materials with close-packed local structures, like metallic glasses, remains to be clarified. Our results indicate that, for the first coordination shell of these systems, the fine radial basis of SOAP does not reveal qualitatively new geometrical features, compared with the much simpler SBO descriptor. We also found that even the first few PC directions are sensitive to the choice of hyper-parameters and also to thermal noise in the particle configurations.[59] We expect that studying the mixtures at much lower temperature and combining separate descriptors for neighbors of different species, as done in Ref. 24, will provide more insight into this issue.

IV. CORRELATIONS OF STRUCTURE WITH DYNAMICS

Having identified a few collective variables that capture structural heterogeneity, one obvious question is whether they are related to *dynamic* heterogeneity. To address this point, we follow the usual approach of measuring the particles' square displacements in the iso-configurational ensemble.[60] Given an initial equilibrium configuration sampled at a temperature T, the propensity for motion of particle i after a time interval t is

$$\mu_i(t) = \langle |\boldsymbol{r}_i(t) - \boldsymbol{r}_i(0)|^2 \rangle_{ic},$$

where $\langle \dots \rangle_{ic}$ denotes an average over an ensemble of independent trajectories that start from the initial positions $\{\boldsymbol{r}_i(0)\}_{i=1,\dots,N}$ with initial velocities drawn from the Maxwell–Boltzmann distribution at temperature T. Hereafter, we refer to the propensity for motion simply as the propensity. Data for the propensity in this section were obtained in Ref. 21, to which we direct the reader for further details.

Spatial fluctuations of the propensity are directly connected to the structure of the initial configuration. Thus, by construction, the propensity captures the structural component of dynamic heterogeneities.[61] The key issue is then to identify *which aspects* of structure correlate with the propensity. This challenge is closely related to the task of predicting localized plastic events in sheared amorphous solids using only structural data.[62]

A. On the relationship between structural and dynamic heterogeneity: Insights from PCA

To assess this, we computed the absolute value of the Pearson correlation coefficient between the propensity and each of the reduced features obtained from PCA. Our analysis focuses on two reference models: the KA and Wahn mixtures. Since the small and big particles of these systems have quite different dynamics, the calculation is carried out separately for each species, yielding distinct correlation coefficients R_A and R_B for big and small particles, respectively. We emphasize that this correlation is computed at the single-particle level as a correlation between μ_i and $\widetilde{X}_j$, contrary to the majority of recent studies on this topic where the structural descriptor is averaged over some length scale.[13,21,39,52,63] Since there is no physically motivated choice for this length scale yet, our approach allows a simple comparison between different structural observables, even though larger absolute values for the correlation coefficient would be obtained if local averaging were performed. This increase is particularly strong at long times, due to a coupling with the local density.[64]

We start by analyzing the small particles, on which our analysis has focused so far. In the top panels of Fig. 10, we show the absolute values of the correlation coefficient, $|R_B|$, between the propensity at time t and each of the reduced features $\widetilde{X}_j$ obtained from the PCA of the SBO descriptor. The first PC gives the strongest correlation. For both models, the correlation displays a flat maximum and starts decaying around the structural relaxation time τ_α, which is measured from the decay of the total self-intermediate scattering function.[21] The values of τ_α are 1.6×10^2 and 9×10^2 for the Wahn and KA mixtures, respectively. The maximum of the correlation coefficient is around 0.5 in the Wahn mixture. For the KA mixture, the first PC still carries significant information about the dynamics, but the correlation is weaker (max. ≈ 0.35). In both models, these results show that a single structural variable, $\widetilde{X}_1$, obtained without any supervision, has a significant correlation with the dynamics of the small particles. Comparing the two models, correlations between structure and dynamics are stronger in the Wahn mixture, consistent with previous work.[13,21] For the higher PCs, individual correlations with dynamics are weaker than PC$_1$.

Corresponding results for big particles are shown in the bottom panels of Fig. 10. Compared to the small particles, the correlations are much weaker and there is no clear separation between the first

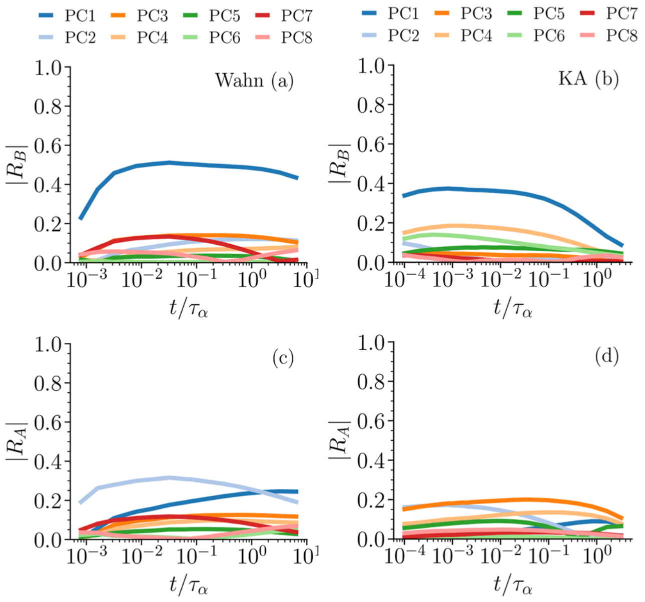

FIG. 10. Pearson correlation coefficients between the propensity at time t and each of the reduced feature $\widetilde{X}_j$ obtained from the PCA of the SBO descriptor, for the Wahn mixture [(a) and (c)] and the KA mixture [(b) and (d)]. Results are shown separately for the small particles [$|R_B|$, (a) and (b)] and the big particles [$|R_A|$, (c) and (d)].

PC and the rest. This points to a more complicated relationship between structure and dynamics, in this case.

B. Multiple linear regression of the propensity of motion using structural data

To characterize more complex structure–dynamics relationships, one must go beyond correlations between individual PCs and propensity. Supervised learning methods provide a data-driven route to identify such relationships in supercooled liquids and they have been used quite intensively in the recent years.[39,65–68] In the absence of a physically motivated framework for such an analysis, we follow the simplest approach and perform a multiple linear regression (LR) to connect the propensity to structural data.[39]

Consider a generic set of P structural features: $X_1, \ldots, X_P$. We least-square fitted the propensity after time t to a linear combination of structural features by defining, for each particle i,

$$\mu_i^X(t; \beta_1, \ldots, \beta_P) = \beta_1 X_1(i) + \cdots + \beta_P X_P(i), \tag{26}$$

where $\beta_1, \ldots, \beta_P$ are fitting parameters. The optimal parameters $\hat{\beta}_1, \ldots, \hat{\beta}_P$ are determined by minimizing the cost function

$$\chi = \frac{1}{N_\alpha} \sum_{i=1}^{N_\alpha} |\mu_i^X(t; \beta_1, \ldots, \beta_P) - \mu_i(t)|^2, \tag{27}$$

where the sum is restricted to all the particles of species α in the sample. The fit was performed using the `LinearRegression` function of the `scikit-learn` package.[69] We also tested the effect of weight regularization in the cost function (the Ridge method[42]), as done in Ref. 39, but this did not lead to any significant change in the results.

For each particle i, the fit yields an interpolated value of the propensity $\hat{\mu}_i = \mu_i^X(t; \hat{\beta}_1, \ldots, \hat{\beta}_P)$, which can be interpreted as a data-driven prediction for the propensity, based on the structure. To assess the quality of these predictions, we use Pearson correlation coefficients R between the interpolated and the actual propensity, which can be obtained by normalizing and shifting the optimal cost χ.

We performed this test for three different structural descriptors. As a reference, we first consider the full SBO descriptor: $X_j = X_j^{\text{SBO}}$ with $j = 1$–8. Second, we consider principal component regression (PCR),[42] in which we fit a linear combination of $\widetilde{X}_1^{\text{SBO}}$ and $\widetilde{X}_2^{\text{SBO}}$. Third, we considered the physically motivated structural measures introduced in Sec. III C: $X_1 = z$, $X_2 = \ell$, $X_3 = \Theta$, $X_4 = u$ (coordination, LFS membership, tetrahedrality parameter, local potential energy).

Results are shown in Fig. 11. For small particles, very similar results are obtained for all three descriptors. We conclude that the two most relevant PCs capture almost all of the correlations between the SBO descriptor and the dynamics and that these descriptors predict the dynamics just as well as the structural measurements. That is, two PCs are enough, in these cases, to capture the bulk of the correlations. For big particles (lower panels of Fig. 11), the correlations are weaker, as observed for individual PCs in Fig. 10. There are larger differences between the LR results for full SBO and the PCR, indicating that the structure–dynamics correlation is

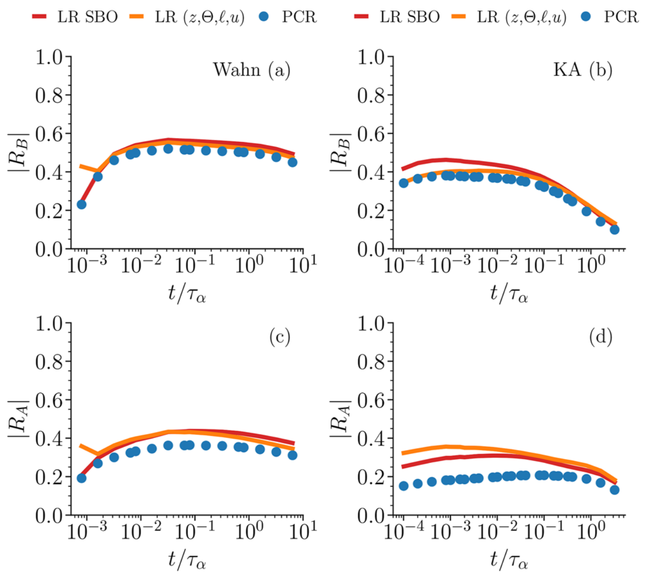

FIG. 11. Pearson correlation coefficients between the propensity at time t and a multiple linear regression of different structural measures, for the Wahn mixture [(a) and (c)] and the KA mixture [(b) and (d)]. Results are shown separately for the small particles [$|R_B|$, (a) and (b)] and the big particles [$|R_A|$, (c) and (d)]. The three sets of features used for fitting are described in the main text.

spread over more than two PCs. For the Wahn mixture, the SBO correlation is similar to that of the physically motivated structural correlators, while the KA mixture shows that the SBO correlation is somewhat weaker. In this latter case, it appears that bond order misses some aspects of local structure that are important for dynamics.

Overall, the picture that emerges from Figs. 10 and 11 is that the small-particle dynamics have significant correlations with local bond order. Hence, this aspect of structure has significant predictive power for dynamics, which can be captured through one or two PCs. For the big particles, the correlations are weaker and are spread over more PCs. In all cases, the correlations for the KA mixture are weaker than those of the Wahn mixture, consistent with earlier work.[13]

C. Fitting the propensity of motion using extended structural descriptors: SOAP and RBO

We now resolve the local structure in more detail, using the extended descriptors defined in Secs. II A 3 and II A 4. Adding more structural information in this way will reduce the residual cost χ, corresponding to an improved fit of the dynamics. We start with the SOAP descriptor focusing on the first coordination shell, choosing r_{cut} close to the first minimum of the total $g(r)$ and using $n_{\max} = 6$ radial components, giving 189 independent structural features in total. Here, we only fit the full descriptor, without any dimensionality reduction, because we found that the results of PCA for this descriptor are sensitive to the choice of the hyper-parameters and are anyway difficult to interpret. We do not expect any overfitting within our simple linear regression model as the number of datapoints greatly exceeds the number of parameters to fit.

From Fig. 12, we see that the SOAP descriptor does indeed lead to stronger correlations with the dynamics compared to the SBO descriptor. This is especially true for short times in the Wahn mixture and longer times in the KA mixture. The values of the correlation coefficients are also in line with those obtained in supervised learning studies,[68] using support vector machine,[66] graph neural networks,[67] or linear regression of the coarse-grained RBO descriptor with an L2 regularization.[39] The improvement over the SBO descriptor, however, comes at the expense of a lack of interpretability of the results: SOAP involves a much larger number of features and it is difficult to understand which are responsible for the correlation with the dynamics. At any rate, these results indicate that subtle features of the local density within the first coordination shell are responsible for a part of the dynamic fluctuations within the iso-configurational ensemble.

Finally, to capture structural correlations beyond the first coordination shell, we employed the SOAP and RBO descriptors—see Eq. (19)—using a larger cutoff distance r_{cut}. For the SOAP descriptor, we used $r_{\text{cut}} = 2.2$ and $n_{\max} = 8$, to maintain approximately the same radial resolution, yielding 324 structural features. For the RBO descriptor, we used Gaussians of width 0.2, centered on a grid $\{d_n\}$ from $r = 0.9$ to 2.2 with spacing 0.1, yielding 104 structural features. From Fig. 12, we see that extending the descriptors to include the second coordination shell increases slightly the correlations with the propensity. We do not observe significant differences in the correlations obtained using the SOAP and RBO either, the former leading to marginally higher correlations than the latter. This indicates that the additional information retained by the SOAP descriptor, which accounts for couplings between different radial orders in the power spectrum (see Sec. II A 3), is not useful for fitting the propensity. The simpler RBO descriptor seems to retain all the necessary information that can be exploited to fit the dynamics within a simple linear regression model. Overall, our results show that including structural correlation beyond the first coordination shell increases the correlation with the dynamics only slightly, and the origin of the increase is difficult to pinpoint.

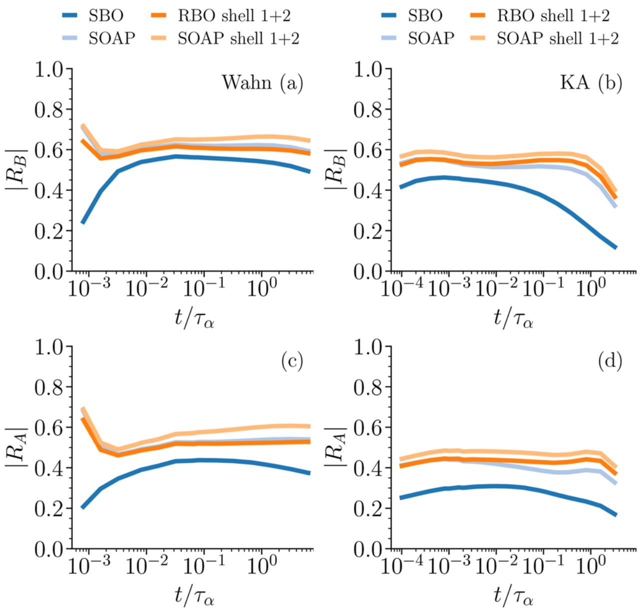

FIG. 12. Pearson correlation coefficients between the propensity at time t and the linear regression of SOAP (blue) and RBO (red) for the Wahn mixture [(a) and (c)] and the KA mixture [(b) and (d)]. Results are shown separately for the small particles [$|R_B|$, (a) and (b)] and the big particles [$|R_A|$, (c) and (d)]. The light colors are for the first coordination shell only, the dark ones up to the second shell.

V. COMPARISON OF DIMENSIONALITY REDUCTION BY PCA AND NEURAL NETWORK AE

Our analysis so far used PCA, which is possibly the simplest dimensionality reduction method. As emphasized in Sec. II B, nonlinear reduction methods provide superior performance when it comes to mapping complex high-dimensional data on a low-dimensional manifold. However, this is often achieved at the cost of a lack of interpretability or the need to fine-tune a large number of free parameters. For instance, the exploration of parameter space and the training process of neural network AEs can quickly become slow and tedious, and the results may not easily generalize. Moreover, while the explainability of machine learning methods recently started to emerge as a separate research field to tackle the problem of the interpretation of their results,[70,71] neural networks are still generally regarded as black boxes.

In this section, we directly compare the results obtained using PCA and the neural network AE, using the SBO descriptor for both the Wahn and KA mixtures. Our model for the AE was developed using `TensorFlow`,[72] following Boattini *et al.*[20] for the choice of free parameters, since the descriptor and the studied physical systems are similar. The neural network is composed of three hidden layers with dimensions $(5M, P, 5M)$, where M is the number of features in the SBO descriptor and P is the dimension of the bottleneck that we set to 2. The input and output layers thus have a dimension of M. Activation functions are the hyperbolic tangent for the encoding and decoding layers and a linear function for the bottleneck and output layers. We use mini-batch stochastic gradient descent with momentum, with a batch size of 200 and a momentum of 0.2, and the learning rate is set to $\alpha = 5 \times 10^{-2}$. We use the mean squared error as loss function, with an L2 penalty term $\lambda = 10^{-5}$ on the weights for regularization. Weights and biases are initialized using the Glorot method.[73] The network is then trained for 250 iterations ("epochs" in the data science jargon) for the Wahn mixture and 500 iterations for the KA mixture on 70% of the dataset, while the remaining 30% are used to evaluate the loss function at the end of each iteration. See Ref. 46 for full details about the parametrization of the network.

For both models, we process the full datasets $\mathbf{X}$ using both PCA and the AE in order to reduce their dimension to $P = 2$ and obtain two datasets $\widetilde{\mathbf{X}}^{\mathrm{PCA}}$ and $\widetilde{\mathbf{X}}^{\mathrm{AE}}$. Since there is no constraint on the range and directions of variance of the data in the reduced space found by the AE, we transform $\widetilde{\mathbf{X}}^{\mathrm{AE}}$ by rotating the reduced descriptors. In terms of the data matrix, this corresponds to

$$\widetilde{\mathbf{X}}_T^{\mathrm{AE}} = \widetilde{\mathbf{X}}^{\mathrm{AE}} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos\theta \end{pmatrix}, \tag{28}$$

which yields a new data matrix $\widetilde{\mathbf{X}}_T^{\mathrm{AE}}$ (the subscript T indicates the additional transformation). The angle of rotation θ is chosen so as to rectify the alignment of the reduced features yielded by the AE to match with those from PCA. In practice, we iteratively rotated $\widetilde{\mathbf{X}}_T^{\mathrm{AE}}$ by steps of $\theta = 5°$ to maximize

$$S_R = |R(\widetilde{X}_{T,1}^{\mathrm{AE}}, \widetilde{X}_1^{\mathrm{PCA}})| + |R(\widetilde{X}_{T,2}^{\mathrm{AE}}, \widetilde{X}_2^{\mathrm{PCA}})|, \tag{29}$$

where $R(\widetilde{X}_{T,j}^{\mathrm{AE}}, \widetilde{X}_j^{\mathrm{PCA}})$ is the Pearson correlation coefficient between the j-th feature of each dataset. This additional transformation on $\widetilde{\mathbf{X}}^{\mathrm{AE}}$ does not affect qualitatively the output distribution of the AE but facilitates the comparison with $\widetilde{\mathbf{X}}^{\mathrm{PCA}}$.

To illustrate this process, Fig. 13 shows an example of the probability densities of the reduced descriptor obtained by the AE for the Wahn mixture, before and after rotation of the dataset. We see that the distribution yielded by the AE is qualitatively similar to those shown for PCA in Fig. 2(c). The only significant difference is a linear transformation (scale and shift), as shown in Fig. 14. This already suggests that the AE does not identify any additional source of structural heterogeneity, beyond what is found in PCA.

Since the training process of the AE has a stochastic nature, we repeated it ten times. We identify the "best" solution by minimizing the loss on the validation set, among the ten repetitions. For each repetition, we also keep track of the EVR and the value of the loss on the validation set at the end of each training, allowing us to assess the variability and robustness of the results. Table I (a and b) shows the absolute value of the average Pearson correlation coefficient between different pairs of features in the reduced descriptors, as well as the average EVR and validation loss, for small and big particles, respectively.

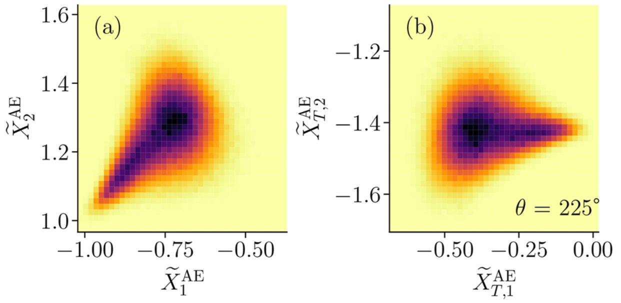

FIG. 13. (a) Probability density $p(\widetilde{X}_1^{\mathrm{AE}}, \widetilde{X}_2^{\mathrm{AE}})$ from a representative solution obtained by the AE with the SBO descriptor in the Wahn model (small particles). (b) Probability density $p(\widetilde{X}_{T,1}^{\mathrm{AE}}, \widetilde{X}_{T,2}^{\mathrm{AE}})$ after a rotation of the reduced descriptor by an angle $\theta = 225°$, which maximizes S_R.

For all the models and types of particles, the variability is extremely low, as shown by the small standard deviation on the correlation. Remarkably, the correlations between "related" features $(\widetilde{X}_{T,j}^{\mathrm{AE}}, \widetilde{X}_j^{\mathrm{PCA}})$ are very close to 1, while the correlations between "opposite" features $(\widetilde{X}_{T,j}^{\mathrm{AE}}, \widetilde{X}_k^{\mathrm{PCA}}, j \neq k)$ are close to 0. Moreover, the average EVR of the AE is identical to that of PCA, which shows that the AE and PCA effectively keep the same amount of information from the original data. We obtained very similar results for the bare BO descriptor (not shown). There is almost no difference between the validation losses of the various training iterations, which suggests that they all approach the same minimum of the loss function. We note that the validation loss is largest for the small particles of the KA mixture, possibly because the number of such particles is 2.5 times lower than the corresponding number for the Wahn mixture.

In conclusion, we found that the output of PCA and AE is almost identical, supporting the idea that the bond-order descriptors

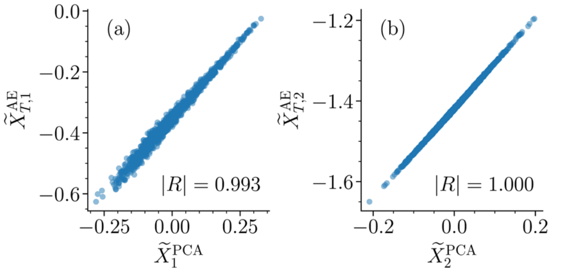

FIG. 14. Scatter plots of $\widetilde{X}_j^{\mathrm{PCA}}$ against the reduced features from the AE after rotation, $\widetilde{X}_{T,j}^{\mathrm{AE}}$ for (a) $j = 1$ and (b) $j = 2$. The distribution considered for the AE is the same as in Fig. 13. Apart from a shift, the two distributions are almost identical, as confirmed by the Pearson correlation coefficients $|R|$. Results are shown for the SBO descriptor in the Wahn model (small particles). For visualization purposes, only a subset of 1000 datapoints are shown.

TABLE I. Comparison between the reduced descriptor yielded by PCA, ($\widetilde{X}_1^{\rm PCA}, \widetilde{X}_2^{\rm PCA}$), and the reduced descriptor yielded by the AE, ($\widetilde{X}_{T,2}^{\rm AE}, \widetilde{X}_{T,2}^{\rm AE}$), rotated so as to maximize S_R. The first two rows show the absolute value of the Pearson correlation coefficient between the first and second features of both methods, the third row shows the total EVR, and the fourth row shows the loss on the validation set of the AE at the end of the training procedure. Errors are the standard deviation of each quantity over ten repetitions of the AE.

(a) Small particles				
	Wahn		KA	
	$\widetilde{X}_1^{\rm PCA}$	$\widetilde{X}_2^{\rm PCA}$	$\widetilde{X}_1^{\rm PCA}$	$\widetilde{X}_2^{\rm PCA}$
$\widetilde{X}_{T,1}^{\rm AE}$	0.998 ± 0.003	0.03 ± 0.02	0.998 ± 0.003	0.05 ± 0.03
$\widetilde{X}_{T,2}^{\rm AE}$	0.06 ± 0.04	0.999 ± 0.001	0.05 ± 0.04	0.998 ± 0.003
EVR	0.801 6 ± 0.000 6 (0.800 9 for PCA)		0.680 ± 0.001 (0.681 for PCA)	
Validation loss	0.000 73 ± 0.000 01		0.001 01 ± 0.000 01	
(b) Big particles				
	Wahn		KA	
	$\widetilde{X}_1^{\rm PCA}$	$\widetilde{X}_2^{\rm PCA}$	$\widetilde{X}_1^{\rm PCA}$	$\widetilde{X}_2^{\rm PCA}$
$\widetilde{X}_{T,1}^{\rm AE}$	0.999 ± 0.001	0.03 ± 0.01	1.000 ± 0.001	0.03 ± 0.02
$\widetilde{X}_{T,2}^{\rm AE}$	0.05 ± 0.02	0.999 ± 0.001	0.02 ± 0.01	0.999 ± 0.001
EVR	0.750 1 ± 0.000 6 (0.750 3 for PCA)		0.699 3± 0.000 8 (0.699 5 for PCA)	
Validation loss	0.000 594 ± 0.000 009		0.000 551± 0.000 004	

do not have a complex distribution that would benefit from a nonlinear dimensionality reduction method. Indeed, the explained variance ratio of the AE is not larger than that of PCA, suggesting that it does not capture any additional information compared to PCA, at least in dimension $P = 2$ where the majority of the variance is already restored by both methods. It appears that PCA is able to capture the key structural features just as accurately as the AE, both in systems with strong and weak structural heterogeneity. In the context of this study on simple binary mixtures, it appears that the reward of using a complex machine learning method such as the AE does not justify the difficulty in setting it up. This conclusion is broadly consistent with the findings of recent work on supervised learning of glassy dynamics.[39,68]

VI. DISCUSSION AND PERSPECTIVES

We close with a broad view of the context of these results and open problems left for future studies.

A. Toward a robust operational definition of locally favored structures

As noted in the Introduction, a central motivation for unsupervised learning in glassy binary mixtures is to pursue a robust definition of locally favored structures, without reference to dynamic properties nor prior knowledge of the relevant symmetries of the particles' arrangements. Conventional classification methods[5–9] yield a large discrete set of distinct local structures, but many of these are easily transformed into one another by thermal distortions. Our analysis focused instead on dimensionality reduction of smooth high-dimensional descriptors, as a step toward a coarse-grained notion of LFS.

Among the systems studied, the Wahn and $Cu_{36}Zr_{64}$ models have pronounced icosahedral order at low temperatures, which has a clear signature in the probability density of the reduced descriptors. This allows identification of the LFS, as a distinct "mode" in the distribution of the reduced descriptor. Indeed, one might use the PC_1 defined here as a smooth measure of icosahedral order in these systems. Similarly, the SiO_2 model has directional bonding, which leads to an heterogeneous feature space in which the preferred tetrahedral arrangements are clearly distinguished from other defective structures. For other systems, i.e., the KA mixture and $Ni_{33}Y_{67}$ model, the structural feature space appears rather homogeneous and the putative LFS, determined from the largest occurrence of cell signatures in the Voronoi tessellation,[10] does not stand out from the background of irregular local structures. Given the range of dimensionality reduction methods and descriptors that we tried, we suspect that this is not a technical limitation but rather an intrinsic property of these amorphous systems. Note that the results may be qualitatively different with descriptors that treat separately particles of different species,[16] which provide a richer description of compositional fluctuations.[24]

Overall, it seems that some glassy mixtures do have well-defined LFS, while others continue to explore a broad spectrum of disordered local structures, even at relatively low temperatures. The latter situation is assumed in mean-field theories of the glass transition,[74,75] while the first is central to structure-based approaches.[1,76] It would be interesting to extend this analysis to more complex glassy systems, like polymers and organic glasses, to understand the extent to which LFSs appear there and whether they can be detected by unsupervised methods. Finally, a robust operational definition of LFS requires careful consideration of two additional aspects, *viz.*, modality and temperature dependence of the feature space, to which we turn in Sec. VI B.

B. Modality and temperature dependence of structural heterogeneity

While the results for the glassy models with strong icosahedral order in Fig. 2 indicate a population of LFS that is distinct from the rest of the liquid, it would be interesting to understand more precisely if the higher-dimensional structural descriptor can be decomposed into a large number of structural states (or modes). A related problem arises in studies of polydisperse colloidal suspensions, when looking for an effective description of structure and thermodynamics in terms of a discrete number of families of particles.[77,78] Identifying the number of distinct modes of a high-dimensional distribution is a delicate task in unsupervised learning[42] and even well-established tests used in cluster analysis (such as the Bayesian information criterion, used for instance in Ref. 20) may fail under some conditions.[79]

In this work, we have avoided the question of *how many* modes best fit the distribution of the structural descriptors, but we did attempt a global assessment of the modality of the distribution $p(\widetilde{X})$, to tell whether it is better described by one or multiple modes. Specifically, we tested a recently proposed statistical criterion called the folding test of unimodality,[79] which can also be applied to high-dimensional datasets. The test outputs a nonnegative folding ratio parameter Φ, which is normalized such that $\Phi < 1$ for multimodal distributions. The normalization considers the approximate folding statistics of the uniform random vector in dimension d,[79] but we found that this normalization leads to a trivial scaling with the dimensionality of our datasets, which makes it unreliable. It would be interesting to test a suitably modified version of this metric in a future study.

When distinguishing systems that display a well-defined LFS from those with a uniform spectrum of disordered structures, it is important to consider the role of temperature. One might expect a broad range of structures at high temperature, with LFSs emerging on cooling. Previous work[21] suggested different degrees of structural fragility in the KA and Wahn model, in that the structural heterogeneity of Wahn has a stronger temperature dependence. We performed a preliminary assessment of the temperature evolution of the EVR, the variance of the PCs, and the non-Gaussianity of their distributions. The results (not shown here) broadly support the idea that the models with icosahedral order studied here have a higher structural fragility than the others. Finally, we note that while the current unsupervised learning approach can be used to identify structural modes at any given temperature, it is unclear how the corresponding directions in the high-dimensional feature space evolve with temperature. This should be taken into account by a robust and transparent definition of LFS.

C. Beyond the first coordination shell: How heterogeneous is medium range order?

In most of the previous sections, we probed structural heterogeneity over a length scale corresponding to the first coordination shell, i.e., we focused on local order. However, in the analysis carried in Sec. IV C, we extended the range of the SOAP descriptor to include the second coordination shell. The results shown in Fig. 12 indicate that this brings only a marginal contribution to the structure–dynamics relationship, hence the dominant source of structural heterogeneity comes from the first coordination shell.

A recent study by Zhang and Kob showed that bond order persists even at large distances in liquids, both under normal and supercooled conditions,[80] suggesting that there is interesting structural heterogeneity beyond the first shell. For instance, the KA mixture is characterized by alternating icosahedral and dodecahedral order in successive coordination shells around big particles. To characterize the heterogeneity of such order beyond the first coordination shell, we analyzed the local structure of the Wahn mixture using the RBO descriptor, restricting ourselves this time only to the second coordination shell. We used the following two grids of distances for the descriptor: (i) $\{d_n\} = (0.9, 1.0, 1.1, 1.2, 1.3)$ and (ii) $\{d_n\} = (1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2)$, using Gaussians of width $\delta = 0.2$. We found that the probability densities $p(\widetilde{X}_1, \widetilde{X}_2)$ changed from bimodal to unimodal, when switching from (i) to (ii), indicating either that our current methodology is not sufficient to capture the heterogeneities in this longer-ranged order or perhaps that there is little heterogeneity on such scales, even if there is significant ordering. Similarly, the distributions of the reduced descriptors that characterize the second coordination shell appeared unimodal in the KA mixture. We defer to a future study a more in-depth analysis of higher order coordination shells, in both close-packed and network-forming systems, as well as of the temperature dependence of the heterogeneity of medium range order.

D. A critical assessment of unsupervised learning methods for glassy systems

The words “unsupervised machine learning” might suggest a procedure where the user passes some input configurations to an algorithm, which automatically returns a characterization of the local structure, without requiring any assumption as to the dominant structural motifs. However, this picture is too simplistic because it neglects at least two places in which the user’s assumptions are baked into the method: the choice of the descriptor and of its hyperparameters. We have emphasized throughout this work that the results depend on these choices: It is essential to keep in mind that they can strongly influence the results.

To illustrate this, we briefly mention two pitfalls that we encountered in this study. First, it may seem natural to use descriptors that include as much information as possible to avoid biasing the algorithm toward any particular form of order. However, in the current context, this hinders interpretability of the results, making it necessary to relate *a posteriori* the reduced variables to other physically motivated quantities. Moreover, a high-dimensional descriptor such as SOAP contains much more information than the simple SBO descriptor, but the dimensionality reduction yields qualitatively similar structural features in both cases, compare Figs. 2 and 7. The reduced descriptors extracted from SOAP also depend sensitively on its hyper-parameters, see Fig. 9, whose choice depend on the identification of the relevant length scales in the system.

The second pitfall is that a bad choice of descriptor can lead to detection of erroneous structure, even in a disordered system. This can be appreciated from Fig. 3, which appears to show complex distributions of the first two PCs in all the models considered, hinting at the existence of different local structures. In fact, these structures appear because the descriptor detects neighboring particles according to a fixed cutoff, which splits a continuous family of local structures into distinct subpopulations. A more reliable representation of the data is obtained by using a smoothed cutoff (Fig. 2), which reveals the underlying continuous family.

Based on these observations, our conclusion is that unsupervised learning provides a powerful tool in this context, but its application and interpretation still requires care, as with all methods. It should not be surprising that an algorithm requires some guidance to identify the aspects of local structure that are physically relevant in glasses. Providing appropriate guidance requires in turn some physical insight from the user as well as an understanding of how the data will be used within the algorithm.

ACKNOWLEDGMENTS

We thank A. Rodriguez for useful discussions. Part of the simulations were carried out on the CINECA HPC cluster within the CINECA-University of Trieste agreement.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Daniele Coslovich: Writing – review & editing (equal). **Robert L. Jack**: Writing – review & editing (equal). **Joris Paret**: Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.7108317.

APPENDIX A: MODELS

1. Wahn mixture

The Wahn model, introduced by Wahnström in Ref. 26, is a two-component Lennard-Jones mixture composed of type-A (big) and type-B (small) particles with chemical fractions $x_A = x_B = 0.5$. Particles interact via a Lennard-Jones potential,

$$u_{\alpha\beta}(r) = 4\epsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r}\right)^{6}\right], \tag{A1}$$

where α and β are species indices. The values for the interaction parameters are given in the original paper.[26] Quantities are expressed in the following system of units: The unit of length is σ_{AA}, the unit of energy is ϵ_{AA}, and the unit of time is $\sqrt{m_A\sigma_{AA}^2/\epsilon_{AA}}$, this is also valid for the other models presented in the next paragraphs except for the one in Appendix A 5. The number density is $\rho = N/V = 1.297$, where V is the volume of the cubic simulation cell. We consider molecular dynamics simulation data, produced in the context of Ref. 21, for a system of $N = 20\ 000$ particles. In the following, we analyze $n_{\mathrm{conf}} = 10$ statistically uncorrelated configurations at temperature $T = 0.58$. The estimated mode-coupling crossover temperature is $T_{\mathrm{MCT}} \approx 0.56$.

2. KA mixture

The Kob-Andersen (KA) mixture[25] is loosely designed to reproduce the structure of the $\mathrm{Ni}_{80}\mathrm{P}_{20}$ metallic glass former using Lennard-Jones interactions. It is composed of type-A (big) and type-B (small) particles with chemical fractions $x_A = 0.8$ and $x_B = 0.2$, and the number density is set to $\rho = 1.2$. As for the Wahn mixture, we consider molecular dynamics simulation data, produced in the context of Ref. 21, for a system of $N = 20\ 000$ particles. We consider $n_{\mathrm{conf}} = 10$ statistically uncorrelated configurations at temperature $T = 0.45$ ($T_{\mathrm{MCT}} \approx 0.435$).

3. SiO_2 model

We also study a simple binary model, based on short-range interactions, that mimics the structure and dynamics of amorphous silica.[28] The chemical fractions are $x_{\mathrm{Si}} = 0.33$ and $x_{\mathrm{O}} = 0.67$. The interaction potential between Si and O particles is of the Lennard-Jones type, as in Eq. (A1), while the one between identical species is a simple inverse power law,

$$u_{\alpha\alpha} = 4\epsilon_{\alpha\alpha}\left(\frac{\sigma_{\alpha\alpha}}{r}\right)^{12}. \tag{A2}$$

The interaction parameters are given in the original paper. The simulation data analyzed in this work were obtained in Ref. 30. The number of particles is $N = 2000$. We consider $n_{\mathrm{conf}} = 20$ statistically uncorrelated configurations at temperature $T = 0.3397$ ($T_{\mathrm{MCT}} \approx 0.31$).

4. $\mathrm{Ni}_{33}\mathrm{Y}_{67}$ model

We consider a parametrization of the LJ potential that provides a realistic description of the structure of amorphous alloys of Ni and Y atoms.[27] This model is characterized by a single energy scale and diameters $\sigma_{\alpha\beta}$, determined by fitting structural data on experimental Ni–Y alloys at several compositions. As in Ref. 27, we use chemical fractions $x_{\mathrm{Ni}} = 0.33$ and $x_{\mathrm{Y}} = 0.67$. In the following, we will refer to Ni and Y atoms in this mixture as small and big particles, respectively. We carried out molecular dynamics simulations for a system composed of 4000 particles, which we cooled at constant density $\rho = 1.5$. This corresponds approximately to the highest densities reached along an isobaric path at $P = 10$ in Ref. 10. The system is then equilibrated and simulated at $T = 0.55$ ($T_{\mathrm{MCT}} \approx 0.52$) and we collect $n_{\mathrm{conf}} = 40$ statistically uncorrelated configurations.

5. $\mathrm{Cu}_{64}\mathrm{Zr}_{36}$ model

We simulate with the LAMMPS molecular dynamics code[81] an embedded-atom model for CuZr alloys using the interatomic potentials developed in Ref. 29. We use a model system of size $N = 16\ 000$ with chemical fractions $x_{\mathrm{Cu}} = 0.64$ and $x_{\mathrm{Zr}} = 0.36$, which displays a pronounced icosahedral local order at low temperatures.[82,83] Note that such local order develops prominently around the (small) Cu particles. The model is also prone to crystallization into a $CuMg_2$ Laves phase[84] and we confirmed this kind of instability at zero pressure for temperatures in the range between 800 and 900 K. We could, however, equilibrate and simulate the system for 0.12 ns at zero pressure and temperature $T = 800$ K, while observing only a modest minor drift of the potential energy. At this temperature, the structural relaxation time τ_α is about 3 ps and we consider $n_{\mathrm{conf}} = 11$ statistically uncorrelated configurations. Given that it is difficult to avoid crystallization for temperatures above 800 K, we do not have a precise estimate of the MCT crossover temperature for this model. However, inspection of the data reported in Refs. 82 and 83 suggests that it may lie in the range 750–800 K.

TABLE II. Eigenvectors $V^{(1)}$ and $V^{(2)}$ of the covariance matrix, corresponding to the two largest eigenvalues of the SBO descriptor for the small particles of the close-packed mixtures and for the Si particles in the SiO_2 models.

	EVR	Q_1^S	Q_2^S	Q_3^S	Q_4^S	Q_5^S	Q_6^S	Q_7^S	Q_8^S
KA									
PC_1	37%	+0.06	+0.07	+0.19	+0.66	−0.51	+0.05	+0.49	+0.12
PC_2	31%	+0.07	+0.01	+0.02	−0.39	−0.50	+0.65	+0.01	−0.40
$Ni_{33}Y_{67}$									
PC_1	38%	+0.09	+0.06	+0.06	+0.25	−0.82	+0.31	+0.40	−0.03
PC_2	29%	−0.02	−0.00	+0.10	+0.78	+0.12	−0.41	+0.10	+0.43
Wahn									
PC_1	61%	−0.03	−0.04	−0.06	−0.13	−0.41	+0.68	−0.50	−0.31
PC_2	20%	−0.05	+0.01	+0.03	+0.12	+0.72	+0.09	−0.64	+0.22
$Cu_{64}Zr_{36}$									
PC_1	65%	−0.03	−0.04	−0.05	−0.16	−0.53	+0.65	−0.42	−0.31
PC_2	18%	+0.03	−0.01	−0.01	−0.07	−0.69	−0.16	+0.70	−0.09
SiO_2									
PC_1	65%	+0.25	+0.61	−0.05	+0.04	+0.60	−0.08	−0.02	+0.44
PC_2	16%	+0.11	−0.41	−0.34	+0.29	+0.04	−0.71	+0.17	+0.28

TABLE III. Eigenvectors $V^{(1)}$ and $V^{(2)}$ of the covariance matrix, corresponding to the two largest eigenvalues of the SBO descriptor for the big particles of the close-packed mixtures.

	EVR	Q_1^S	Q_2^S	Q_3^S	Q_4^S	Q_5^S	Q_6^S	Q_7^S	Q_8^S
KA									
PC_1	48%	−0.01	−0.01	+0.00	+0.04	+0.12	+0.75	−0.64	+0.07
PC_2	22%	+0.06	+0.03	+0.03	+0.06	+0.06	−0.43	−0.40	+0.80
$Ni_{33}Y_{67}$									
PC_1	46%	−0.02	−0.02	−0.01	+0.04	+0.19	+0.87	−0.42	−0.12
PC_2	25%	+0.07	+0.02	+0.03	+0.02	−0.01	−0.21	−0.65	+0.72
Wahn									
PC_1	49%	−0.00	−0.00	+0.02	+0.07	+0.20	+0.74	−0.63	−0.06
PC_2	26%	+0.04	+0.01	+0.04	+0.09	+0.22	−0.39	−0.44	+0.77
$Cu_{64}Zr_{36}$									
PC_1	41%	+0.07	+0.04	+0.01	+0.08	+0.24	+0.41	−0.82	+0.31
PC_2	36%	+0.00	+0.02	−0.01	+0.01	+0.07	+0.73	+0.13	−0.67

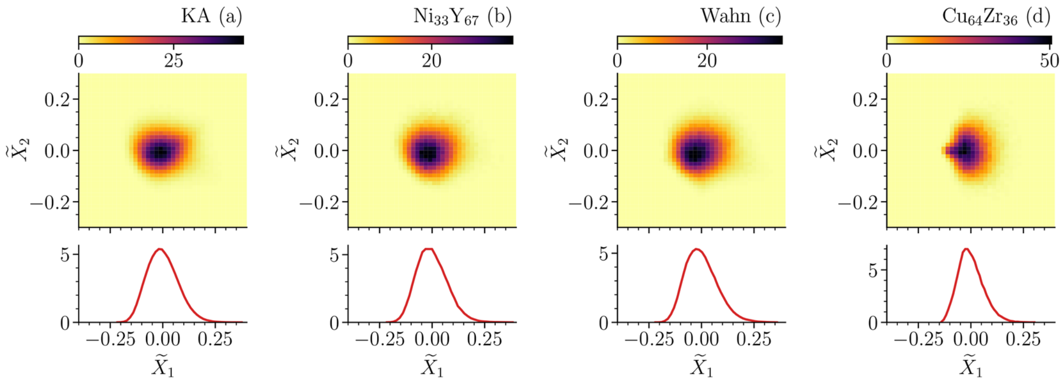

FIG. 15. Same as Fig. 2 but for the large particles of the close-packed mixtures: (a) KA, (b) $\mathrm{Ni_{33}Y_{67}}$, (c) Wahn, (d) $\mathrm{Cu_{64}Zr_{36}}$.

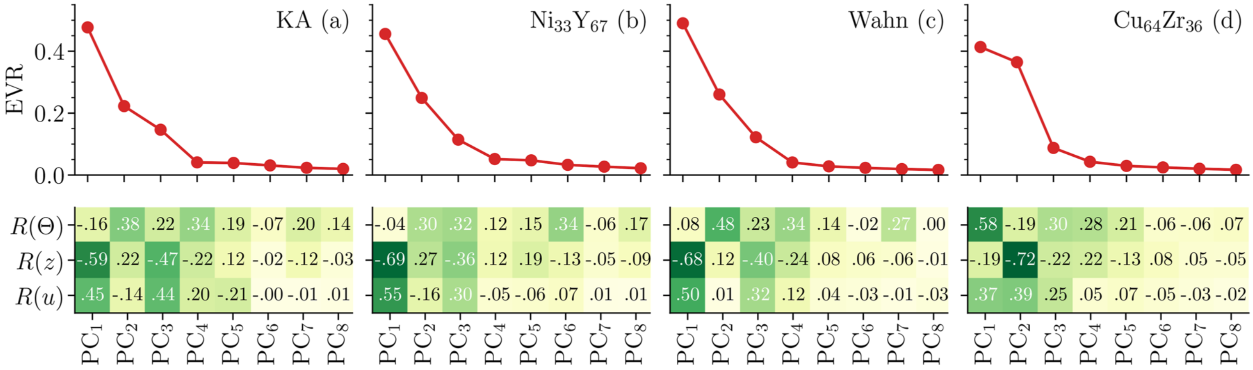

FIG. 16. Same as Fig. 5 but for the large particles of the close-packed mixtures: (a) KA, (b) $\mathrm{Ni_{33}Y_{67}}$, (c) Wahn, (d) $\mathrm{Cu_{64}Zr_{36}}$. Note that no well-defined LFS can be identified from the Voronoi tessellation around these particles; therefore, the correlation with ℓ is not considered.

APPENDIX B: ADDITIONAL RESULTS ON THE SBO DESCRIPTOR

We provide here supplementary information and results on the PCA of the SBO descriptor in the studied models. Tables II and III show the eigenvectors $V^{(1)}$ and $V^{(2)}$, i.e., the first two PC directions. We also show additional results for the big particles of the close-packed mixtures: Fig. 15 presents the distributions $p(\widetilde{X}_1, \widetilde{X}_2)$ obtained from PCA of the SBO descriptor, while Fig. 16 shows the EVR and the correlation of each PC with the physically motivated structural measures introduced in Sec. III C. Note the little bump in the distribution $p(\widetilde{X}_1, \widetilde{X}_2)$ for $\mathrm{Cu_{64}Zr_{36}}$ in Fig. 15(d), which could be due to the presence of a small crystallite in the sample.

REFERENCES

[1] C. P. Royall and S. R. Williams, Phys. Rep. **560**, 1 (2015).
[2] H. Tanaka, H. Tong, R. Shi, and J. Russo, Nat. Rev. Phys. **1**, 333 (2019).
[3] D. Coslovich, J. Chem. Phys. **138**, 12A539 (2013).
[4] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Phys. Rev. B **28**, 784 (1983).
[5] M. Tanemura, Y. Hiwatari, H. Matsuda, T. Ogawa, N. Ogita, and A. Ueda, Prog. Theor. Phys. **58**, 1079 (1977).
[6] B. J. Gellatly and J. L. Finney, J. Non-Cryst. Solids **50**, 313 (1982).
[7] J. D. Honeycutt and H. C. Andersen, J. Phys. Chem. **91**, 4950 (1987).
[8] A. Malins, S. R. Williams, J. Eggers, and C. P. Royall, J. Chem. Phys. **139**, 234506 (2013).
[9] E. A. Lazar, J. Han, and D. J. Srolovitz, Proc. Natl. Acad. Sci. U. S. A. **112**, E5769 (2015).
[10] D. Coslovich and G. Pastore, J. Chem. Phys. **127**, 124504 (2007).
[11] A. Malins, J. Eggers, C. P. Royall, S. R. Williams, and H. Tanaka, J. Chem. Phys. **138**, 12A535 (2013).
[12] A. Malins, J. Eggers, H. Tanaka, and C. P. Royall, Faraday Discuss. **167**, 405 (2014).
[13] G. M. Hocky, D. Coslovich, A. Ikeda, and D. R. Reichman, Phys. Rev. Lett. **113**, 157801 (2014).
[14] F. Arceri, F. P. Landes, L. Berthier, and G. Biroli, "A statistical mechanics perspective on glasses and aging," in *Encyclopedia of Complexity and Systems Science*, edited by R. A. Meyers (Springer Berlin Heidelberg, 2020), pp. 1–68.
[15] P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, Phys. Rep. **810**, 1 (2019).
[16] 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. Csanyi, Acc. Chem. Res. **53**, 1981 (2020).
[17] E. Boattini, M. Dijkstra, and L. Filion, J. Chem. Phys. **151**, 154901 (2019).
[18] R. van Damme, G. M. Coli, R. van Roij, and M. Dijkstra, ACS Nano **14**, 15144 (2020).
[19] S. Becker, E. Devijver, R. Molinier, and N. Jakse, Phys. Rev. E **105**, 045304 (2022).
[20] E. Boattini, S. Marín-Aguilar, S. Mitra, G. Foffi, F. Smallenburg, and L. Filion, Nat. Commun. **11**, 5479 (2020).
[21] J. Paret, R. L. Jack, and D. Coslovich, J. Chem. Phys. **152**, 144502 (2020).
[22] V. L. Deringer, M. A. Caro, R. Jana, A. Aarva, S. R. Elliott, T. Laurila, G. Csányi, and L. Pastewka, Chem. Mater. **30**, 7438 (2018).

[23] B. Monserrat, J. G. Brandenburg, E. A. Engel, and B. Cheng, Nat. Commun. **11**, 5757 (2020).

[24] A. Offei-Danso, A. Hassanali, and A. Rodriguez, J. Chem. Theory Comput. **18**, 3136 (2022).

[25] W. Kob and H. C. Andersen, Phys. Rev. Lett. **73**, 1376 (1994).

[26] G. Wahnström, Phys. Rev. A **44**, 3752 (1991).

[27] R. G. Della Valle, D. Gazzillo, R. Frattini, and G. Pastore, Phys. Rev. B **49**, 12625 (1994).

[28] D. Coslovich and G. Pastore, J. Phys.: Condens. Matter **21**, 285107 (2009).

[29] Y. Q. Cheng, H. W. Sheng, and E. Ma, Phys. Rev. B **78**, 014207 (2008).

[30] L. Berthier, G. Biroli, D. Coslovich, W. Kob, and C. Toninelli, Phys. Rev. E **86**, 031502 (2012).

[31] W. Lechner and C. Dellago, J. Chem. Phys. **129**, 114707 (2008); arXiv:0806.3345.

[32] J. A. van Meel, L. Filion, C. Valeriani, and D. Frenkel, J. Chem. Phys. **136**, 234107 (2012); arXiv:1202.5281.

[33] W. Mickel, S. C. Kapfer, G. E. Schröder-Turk, and K. Mecke, J. Chem. Phys. **138**, 044501 (2013).

[34] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).

[35] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).

[36] B. Parsaeifard, D. S. De, A. S. Christensen, F. A. Faber, E. Kocer, S. De, J. Behler, O. A. von Lilienfeld, and S. Goedecker, Mach. Learn.: Sci. Technol. **2**, 015018 (2021).

[37] L. Himanen, M. O. J. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster, Comput. Phys. Commun. **247**, 106949 (2020).

[38] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Phys. Chem. Chem. Phys. **18**, 13754 (2016).

[39] E. Boattini, F. Smallenburg, and L. Filion, Phys. Rev. Lett. **127**, 088007 (2021).

[40] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, in *Database Theory—ICDT 2001*, edited by J. Van den Bussche and V. Vianu (Springer Berlin Heidelberg, 2001), pp. 420–434.

[41] I. T. Jolliffe and J. Cadima, Philos. Trans. R. Soc., A **374**, 20150202 (2016).

[42] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, New York, 2016).

[43] S. Goldt, M. Mézard, F. Krzakala, and L. Zdeborová, Phys. Rev. X **10**, 041044 (2020).

[44] T. Mendes-Santos, X. Turkeshi, M. Dalmonte, and A. Rodriguez, Phys. Rev. X **11**, 011040 (2021).

[45] J. Paret and D. Coslovich, J. Open Source Software **6**, 3723 (2021).

[46] D. Coslovich, R. L. Jack, and J. Paret (2022). "Dimensionality reduction of local structure in glassy binary mixtures," Zenodo.

[47] H. Tanaka, Eur. Phys. J. E **35**, 113 (2012).

[48] A. Engel and C. Van den Broeck, *Statistical Mechanics of Learning* (Cambridge University Press, Cambridge, 2001).

[49] M. Leocmach and H. Tanaka, Nat. Commun. **3**, 974 (2012).

[50] J. Paret, "Hidden order in amorphous materials," Ph.D. thesis, Université de Montpellier, 2021.

[51] J. A. Hartigan and P. M. Hartigan, Ann. Stat. **13**, 70 (1985).

[52] H. Tong and H. Tanaka, Phys. Rev. X **8**, 011041 (2018).

[53] C. H. Rycroft, Chaos **19**, 041111 (2009).

[54] The range of the components in the reduced SOAP descriptor scales with n [see Eq. (14)], which creates a trivial increase of the variance across features. The features with the largest variance would thus dominate the PCA, hence the need for feature scaling to bring all the features to the same range.

[55] Note that without feature scaling, the PC_1 direction gathers instead a *large* variance, which further increases with increasing σ. Very likely, this reflects the lack of normalization of the descriptor, which gives a strong weight to Q_0, i.e., the local coordination number.

[56] The closest approach distance in our mixtures is typically half of the cutoff distance r_{cut}.

[57] M. O. J. Jäger, E. V. Morooka, F. Federici Canova, L. Himanen, and A. S. Foster, npj Comput. Mater. **4**, 37 (2018).

[58] M. Ceriotti, M. J. Willatt, and G. Csányi, "Machine learning of atomic-scale properties based on physical principles," in *Handbook of Materials Modeling: Methods: Theory and Modeling*, edited by W. Andreoni and S. Yip (Springer International Publishing, Cham, 2018), pp. 1–27.

[59] We found that energy-minimized configurations provide somewhat more robust results upon changes of hyper-parameters.

[60] A. Widmer-Cooper, P. Harrowell, and H. Fynewever, Phys. Rev. Lett. **93**, 135701 (2004).

[61] L. Berthier and R. L. Jack, Phys. Rev. E **76**, 041509 (2007).

[62] 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, Phys. Rev. Mater. **4**, 113609 (2020).

[63] H. Tong and H. Tanaka, Nat. Commun. **10**, 5596 (2019).

[64] R. L. Jack, A. J. Dunleavy, and C. P. Royall, Phys. Rev. Lett. **113**, 095703 (2014).

[65] E. D. Cubuk, S. S. Schoenholz, J. M. Rieser, B. D. Malone, J. Rottler, D. J. Durian, E. Kaxiras, and A. J. Liu, Phys. Rev. Lett. **114**, 108001 (2015).

[66] S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, Nat. Phys. **12**, 469 (2016).

[67] V. Bapst, T. Keck, A. Grabska-Barwińska, C. Donner, E. D. Cubuk, S. S. Schoenholz, A. Obika, A. W. R. Nelson, T. Back, D. Hassabis, and P. Kohli, Nat. Phys. **16**, 448 (2020).

[68] R. M. Alkemade, E. Boattini, L. Filion, and F. Smallenburg, J. Chem. Phys. **156**, 204503 (2022).

[69] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, J. Mach. Learn. Res. **12**, 2825 (2011).

[70] L. Zdeborová, Nat. Phys. **16**, 602 (2020).

[71] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, Entropy **23**, 18 (2021).

[72] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," software available from tensorflow.org, 2015.

[73] X. Glorot and Y. Bengio, in *Volume 9: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, edited by Y. W. Teh and M. Titterington (Proceedings of Machine Learning Research, PMLR, Chia Laguna Resort, Sardinia, Italy, 2010), pp. 249–256.

[74] T. R. Kirkpatrick and D. Thirumalai, "Random first-order phase transition theory of the structural glass transition," in *Structural Glasses and Supercooled Liquids* (John Wiley and Sons, 2012), pp. 223–236.

[75] G. Biroli and J. P. Bouchaud, "The random first-order transition theory of glasses: A critical assessment," in *Structural Glasses and Supercooled Liquids* (John Wiley and Sons, 2012), pp. 31–113.

[76] G. Tarjus, S. A. Kivelson, Z. Nussinov, and P. Viot, J. Phys.: Condens. Matter **17**, R1143 (2005).

[77] M. Ozawa and L. Berthier, J. Chem. Phys. **146**, 014502 (2017).

[78] P. Patel, M. K. Nandi, U. K. Nandi, and S. Maitra Bhattacharyya, J. Chem. Phys. **154**, 034503 (2021).

[79] A. Siffer, P.-A. Fouque, A. Termier, and C. Largouët, in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '18* (Association for Computing Machinery, 2018), pp. 2210–2218.

[80] Z. Zhang and W. Kob, Proc. Natl. Acad. Sci. U. S. A. **117**, 14032 (2020).

[81] S. Plimpton, J. Comput. Phys. **117**, 1 (1995).

[82] R. Soklaski, Z. Nussinov, Z. Markow, K. F. Kelton, and L. Yang, Phys. Rev. B **87**, 184203 (2013).

[83] R. Soklaski, V. Tran, Z. Nussinov, K. F. Kelton, and L. Yang, Philos. Mag. **96**, 1212 (2016).

[84] R. E. Ryltsev, B. A. Klumov, N. M. Chtchelkatchev, and K. Y. Shunyaev, J. Chem. Phys. **145**, 034506 (2016).