UCLA
UCLA Previously Published Works

Title

Revealing hidden medium-range order in amorphous materials using topological data analysis.

Permalink

https://escholarship.org/uc/item/0h4758n9

Journal

Science Advances, 6(37)

Authors

Sørensen, Søren
Biscio, Christophe
Bauchy, Mathieu
et al.

Publication Date

2020-09-01

DOI

10.1126/sciadv.abc2320

Copyright Information

Peer reviewed

Revealing hidden medium-range order in amorphous materials using topological data analysis

Søren S. Sørensen[1], Christophe A. N. Biscio[2], Mathieu Bauchy[3], Lisbeth Fajstrup[2], Morten M. Smedskjaer[1]*

Despite the numerous technological applications of amorphous materials, such as glasses, the understanding of their medium-range order (MRO) structure—and particularly the origin of the first sharp diffraction peak (FSDP) in the structure factor—remains elusive. Here, we use persistent homology, an emergent type of topological data analysis, to understand MRO structure in sodium silicate glasses. To enable this analysis, we introduce a self-consistent categorization of rings with rigorous geometrical definitions of the structural entities. Furthermore, we enable quantitative comparison of the persistence diagrams by computing the cumulative sum of all points weighted by their lifetime. On the basis of these analysis methods, we show that the approach can be used to deconvolute the contributions of various MRO features to the FSDP. More generally, the developed methodology can be applied to analyze and categorize molecular dynamics data and understand MRO structure in any class of amorphous solids.

INTRODUCTION

Characterization and understanding of medium-range order (MRO) structure of amorphous materials continues to be an active area of research (*1*–*4*). In addition to offering new fundamental insights into the glassy state, such research is needed for explaining the variation in various glass properties (*5*). Typically, MRO in covalent glasses is regarded as structural features in the range of 5 to 20 Å, including arrangements of connected polyhedra, superstructural units, and clustering around structural features such as network modifiers (*6*). Neutron and x-ray diffraction techniques have been widely used to characterize MRO structure of glasses by determining the structure factor, $S(Q)$. However, the structural interpretation of $S(Q)$ is nontrivial because of its origin in reciprocal space, and as a result, the origin of the first sharp diffraction peak (FSDP) in $S(Q)$ has been intensively debated. The FSDP has been associated with various anomalies (*6*, *7*), including anomalous temperature (*8*, *9*), pressure (*10*, *11*), and composition behavior (*12*, *13*).

While the relation of the FSDP with MRO structure is widely accepted, numerous structural origins have been assigned to the FSDP. In glassy silicates, the FSDP has been proposed to originate from quasicrystalline and ring-type structures similar to those found in crystalline silica (*3*, *14*). Similarly, for chalcogenide glasses, the FSDP has been assigned to layer-like structures (*8*, *9*) or cluster-like regions with well-defined periodicities (*15*, *16*). The latter interpretation is also related to Elliot's void model, ascribing the anomalous temperature, pressure, and composition behavior of the FSDP to voids around cation-centered clusters explained by a prepeak in the concentration-concentration structure factor [$S_{CC}(Q)$] of the Bhatia-Thornton formalism (*6*). This model has been successfully applied to various amorphous systems and has also explained some of the anomalous behavior of the FSDP (*6*). Besides experimental diffraction measurements, molecular dynamics simulations have deepened

the understanding of the FSDP and MRO structure, owing to their ability to provide real-space structural information. For example, using first-principles molecular dynamics simulations, the interpretation of Elliot has been questioned based on the lacking FSDP in a model GeSe$_2$ liquid, despite the presence of a prepeak in its $S_{CC}(Q)$ (*6*, *17*). Similarly, no experimental prepeak in the $S_{CC}(Q)$ in vitreous SiO$_2$ has been identified based on neutron diffraction experiments (*18*). For alkali silicates, the compositional dependence of the FSDP can be reproduced by classical molecular dynamics simulations, ascribing most of the FSDP anomalies to scattering length differences (*12*, *13*, *19*). More recently, the FSDP in various tetrahedrally coordinated glasses has been ascribed to be a signature of the local ordering of the tetrahedral unit itself (*4*).

Recently, persistent homology, a type of topological data analysis, has been proposed as a tool to describe the MRO structure of SiO$_2$ glass (*20*, *21*), CuZr and Pd$_{40}$Ni$_{40}$P$_{20}$ metallic glasses (*21*, *22*), and amorphous ices (*23*). Persistent homology is a novel approach to analyze high-dimensional datasets, by focusing on the global properties such as data shape and connectivity, with possible applications, e.g., in biology, physics, and chemistry (*24*). In the case of glass structure, atomic trajectories generated from molecular dynamics simulations and the atomic radii are used as input to generate persistence diagrams, which are two-dimensional (2D) histograms of both the fraction and scale of the structural features, most importantly, rings and cavities embedded in the atomic configuration (Fig. 1). As an advantage compared to standard ring size analyses, persistent homology is not limited to chemically bonded structures but provides a fingerprint of the entire arrangement of atoms.

In topological data analysis, and here specifically persistent homology, tools from the mathematical area of algebraic topology are used to extract features such as components, rings, and voids from a shape built from a given network of data points. In the present case, the nodes of the network are the positions of the atoms, which are distributed in the Euclidean space. The procedure for obtaining the persistence diagram of such atomic network is as follows. First, a ball replaces each atom based on its radius. The radius of each ball is then gradually increased with the same increment for all atom types. Following the study by Hiraoka *et al.* (*21*), the increment is an

[1]Department of Chemistry and Bioscience, Aalborg University, DK-9220 Aalborg, Denmark. [2]Department of Mathematical Sciences, Aalborg University, DK-9220 Aalborg, Denmark. [3]Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), Department of Civil and Environmental Engineering, University of California, Los Angeles, Los Angeles, CA 90095, USA.
*Corresponding author. Email: mos@bio.aau.dk

20Na$_2$O-80Si$_2$O glass

Persistence diagram
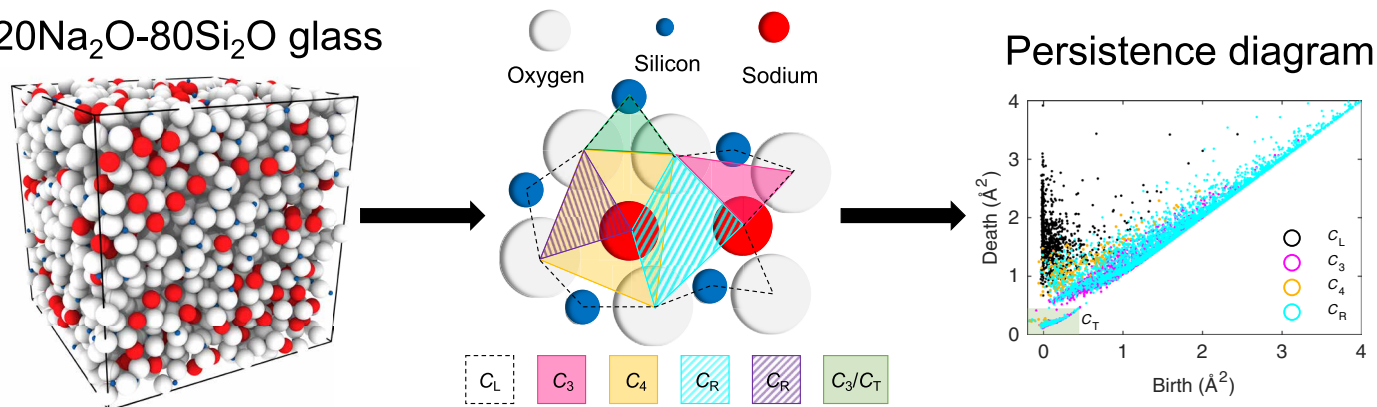
Fig. 1. Schematic overview of the process of obtaining a persistence diagram. Here, the trajectory information for a 20Na$_2$O-80SiO$_2$ glass at 300 K (oxygen, white; silicon, blue; and sodium, red) is used to construct the persistence diagram. The persistence diagram coloring represents the characteristic regions, with the note that the turquoise points represent both the turquoise and violet C_R regions. Stripes indicate that the region is fully or partially overlapping with another region.

increment in the radius squared, and as in the work of Carlsson (*25*) and Biscio and Møller (*26*), we refer to this increment as "time." For calculations, the union of balls is actually described as a union of more complicated convex shapes, which intersect less than the original balls, making computations feasible by giving fewer edges, triangles, and tetrahedra. This construction is equivalent to the union of balls only when the increment of the radius is squared, as described in detail elsewhere (*27*).

This procedure of continuously increasing radii is used to build a sequence of structures with (i) a point for each atom, (ii) an edge between two points when the corresponding growing balls intersect for the first time, (iii) a triangle is filled in when three balls have a common intersection, and (iv) similarly for tetrahedra, etc. The time at which the addition of an edge gives rise to a closed loop is denoted as "birth," while the time at which the addition of a filled triangle gives the first structure bounded by the loop is the "death" of the loop. Similarly, a void is born when the addition of a triangle gives rise to a surface enclosing the void. It dies when a tetrahedron causes the void to be filled. On the other hand, in the simplest case of persistent homology, all connected components are born simultaneously ($b = 0$) and die when an edge between two points intersect for the first time. This gives rise to a persistence diagram, a scatterplot of death versus birth, essentially holding the same information as the radial distribution function (RDF). Such analysis is referred to as 0D persistent homology. Analyses of loops and voids are then denoted as 1D and 2D persistent homology, respectively. In this work, we focus on categorizing loops and voids and use a loop definition based on the first appearance of edges. In the description of the RDF, we rely on the definition of Keen (*28*), while the structure factors are calculated on the basis of the Faber-Ziman formalism (*29*).

In the case of amorphous materials, the loops take the form of geometrical rings of atoms, which are not necessarily covalently bonded. As such, the persistent homology approach differs from traditional ring analysis (*30*). While the latter makes assumptions about the chemical connectivity, the persistent homology approach is unbiased in this sense. Furthermore, while traditional ring analyses are generally meant for categorizing the number of atoms in the rings (a measure also directly obtained by persistent homology), the birth and death of loops conveniently provide quantitative information on both size and shape of the atomic rings.

Generally, the application of persistent homology to amorphous MRO structures is challenging, as the definition of characteristic regions (i.e., groups of loops) adopted in the original studies (*20–23*, *31*) is not unique; that is, they use an "optimal cycle" for the geometric characterizations. Here, we introduce a self-consistent and rigorous definition of characteristic regions by separating them according to the number of contained atoms (see below for definitions). This methodology offers an intuitive understanding of the underlying data (here, glass structure). Moreover, the interpretation of persistence diagrams has previously only been qualitative and relied on visual inspections, making comparisons among different structures highly complex. We propose to overcome these limitations by rigorously categorizing the persistence diagram features according to their structure, e.g., number of atoms and geometry, and by applying the accumulated persistence function (APF) (*26*)

$$\text{APF}(t) = \sum_{i:m_i \leq t}(d_i - b_i) \qquad (1)$$

where b_i and d_i are the birth and death times of point i in the persistence diagram, and $m_i = (b_i + d_i)/2$ is the average length (i.e., "mean age") of point i. With this terminology, a given mean age is equivalent for two points if the points are lying orthogonally to the same point on the diagonal of the persistence diagram. We may also define the "lifetime" of point i as $l_i = d_i - b_i$, which measures how close neighboring atoms are in a loop and weighs it against the separation of the most distant atoms. That is, if local atoms are in proximity, then the loop will be born early and, similarly, if the most distant atoms are in proximity, then the loop will die early. This corresponds to a short lifetime. In other words, to obtain a point with long lifetime, nearby atoms should be close in space but well separated from the most distant atoms.

In detail, the APF is a cumulative sum of all points in the persistence diagram weighted by their lifetime. This means that the APF characterizes the persistence diagram completely and is especially well suited for data where "short-lived" loops or voids (as found in glasses) are real structural features and not noise. The function has previously been applied within the field of spatial statistics (*26*), but here, we apply it to enable direct quantitative comparison of multiple persistence diagrams in a single plot through the shape of the obtained curves. As it will be shown here, the approach is a simple,

yet powerful, alternative to more computationally demanding methods for comparison of persistence diagrams, such as those based on machine learning. Furthermore, to deconvolute the contributions of structural MRO features to the FSDP, we introduce the S_{PH} function, which has previously been used on glassy SiO_2 (21) and two amorphous ices (23)

$$S_{PH,A_k}(Q) = \frac{1}{|A_k|} \sum_{(b\,|\,k,d_i) \in A_k} \delta\left(Q - \frac{2\pi}{l(d_i)}\right) \quad (2)$$

where Q is the scattering vector, $|A_k|$ is the number of elements of a characteristic region (mathematically a subset) A_k, and δ is the Dirac delta distribution. $l(d_i) = 2\sqrt{d_i + r_O^2}$ is the diameter of oxygen atoms (where $r_O = 1.289$ Å is the radius of oxygen) at the time of death of loop i, which is the diameter at which the union of balls fills in the loop, largely corresponding to the "mean" diameter of loop i. This nonintuitive definition of the loop diameter originates from how increments of radius are defined (27). Last, we note that the function of Eq. 2 originates from the definition of a reciprocal space, where a repeating real-space distance, l, is equivalent to a magnitude of $Q = 2\pi/l$ in reciprocal space (so-called Q space). By using the diameter of oxygen at the time of death as the real-space distance descriptor of the loop size, we simply sum up the reciprocal space contributions to S_{PH} of all loops using a summation sign and the Dirac delta distribution. We use the radius of oxygen, as oxygen has the largest radius of the three types of atoms, and because it is the most abundant element in both the glass compositions as well as in the identified loops.

In this work, we use classical molecular dynamics simulations to generate structures of xNa$_2$O-(100 − x)SiO$_2$ ($0 \leq x \leq 40$) glasses (32). By using the persistent homology-based methodology, we categorize MRO structures and their contributions to the FSDP of these sodium silicates, showing good agreement between the FSDP position computed directly from the RDF and that from persistent homology. We chose sodium silicates as a model system because of its well-studied structure (12, 13, 33) and because it is a simple representation of the structural features found in complex silicate glasses, including volumetric compaction and depolymerization of the silicate network upon modifier addition (34). In addition, sodium silicate glasses offer a rich system to investigate the relationship between network topology, persistence diagrams, and the FSDP.

RESULTS

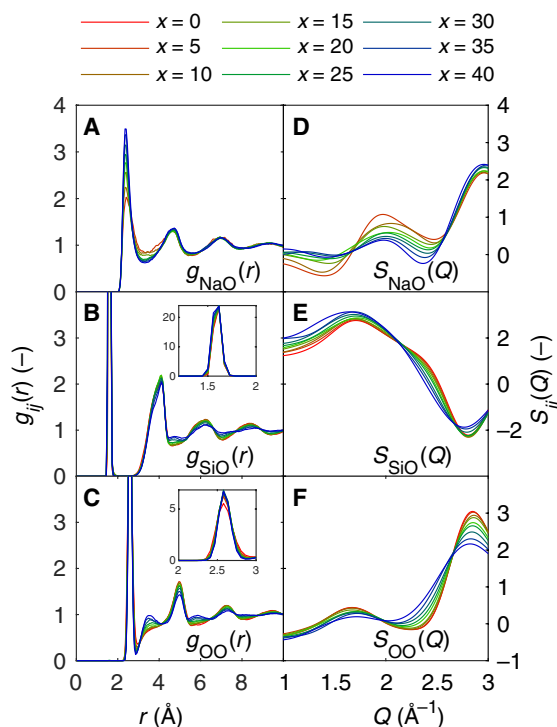### Molecular dynamics simulations of glass structure
As shown in figs. S1 and S2, we generally find good agreement between the simulated and experimental neutron structure factors and Q^n speciation (where Q^n denotes SiO$_4$ tetrahedra with n number of bridging oxygens) (33, 35–38). We note that the minor differences between experiments and simulations do not affect the conclusions of this study. This is because we compare the structural information obtained from the simulated RDF with that from the persistent homology analysis; i.e., both are based on trajectory information from the same molecular dynamics simulations. Generally, the incorporation of Na$_2$O into the silica network causes the fully connected structure to become depolymerized as nonbridging oxygens are formed and internal voids are filled by Na$^+$. The latter is experimentally observed as an increase in the packing factor (34). This is also seen in the present simulations as presented in fig. S3 (39). Similarly,

the ring size distribution of the structure is seen to notably broaden upon increasing Na$_2$O content (fig. S4).

Figure 2 shows the partial RDFs [$g_{ij}(r)$, left column] and partial Faber-Ziman structure factors [$S_{ij}(Q)$, right column] for the (A and D) NaO, (B and E) SiO, and (C and F) OO atomic pairs. The remaining atomic pairs (NaNa, NaSi, and SiSi) are presented in fig. S5. For the Na RDFs, only minor changes are noted upon increasing sodium content, except for a slight decrease of position and intensity of the first peak in the NaNa RDF (fig. S5A) and an increase in the sharpness of the first peak (~2.3 Å) of the NaO RDF (Fig. 2A). For the NaSi (fig. S5B) and SiSi RDFs (fig. S5C), only small changes are found, while the SiO (Fig. 2B) and OO RDFs (Fig. 2C) show the largest variation with increasing sodium content. For SiO, this is noted as a slight decrease in intensity of the second peak (~4.0 Å) and an increase in intensity of a small peak at ~4.8 Å. Similarly, the OO RDF features increasing intensities for the peaks at 2.6, 3.6, and 5.9 Å, while decreasing intensities are noted for peaks at 5.0 and 7.3 Å. The remaining peaks at larger separation of the SiO and OO RDFs exhibit general broadening with increasing Na$_2$O content. Overall, these data suggest that, primarily, the silicate backbone of the glass is affected by the addition of Na$_2$O.

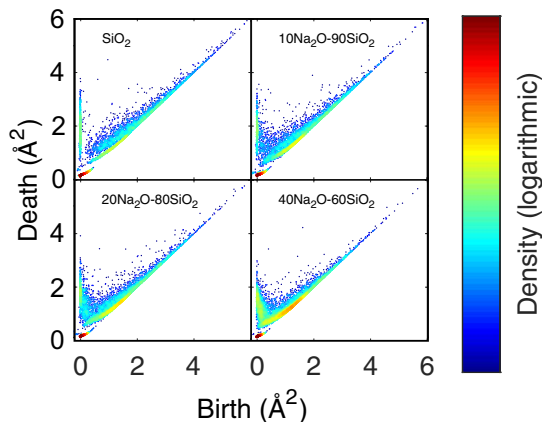### Generation of persistence diagrams
Next, these simulated glass structures are subjected to the topological data analysis. In previous works on persistent homology and glasses (21, 23), nonunique definitions of the characteristic regions with limited structural interpretations were provided. Here, we adopt rigid geometry-based definitions of the presented characteristic regions

Fig. 2. Glass structure analyses from molecular dynamics simulations. Partial RDFs [$g_{ij}(r)$, left column] as well as partial Faber-Ziman structure factors [$S_{ij}(Q)$, right column] for (**A** and **D**) NaO, (**B** and **E**) SiO, and (**C** and **F**) OO atomic pairs. Noticeable composition variation with the Na$_2$O content is observed for all presented correlations. Insets in (B) and (C) show enlargements of the first peak of each $g_{ij}(r)$.

(C). We define C_L as long loops consisting of ≥ 5 atoms, which are found to be large structures of mainly oxygen (67 to 82%), with remaining atoms being both Si and Na. C_3 and C_4 are defined as subloops of C_L, consisting of exactly three and four atoms, respectively. With these definitions, we are left with 1500 to 2500 unaccounted loops (~20% of all loops), which we denote as the remaining three- and four-membered loops (C_R). In addition, we define C_T as the square $[-\infty, 0.45] \times [-\infty, 0.45]$ in the persistence diagram primarily consisting of triangular O-Si-O loops, hence only indicating local structure within the SiO$_4$ tetrahedron. We note that C_T holds loops from both C_3, C_4, and C_R, yet we only use C_T to exclude short-range order points of the persistence diagram in the S_{PH} analysis. The procedure for obtaining a persistence diagram has been illustrated in Fig. 1, which is based on a molecular dynamics simulation of the 20Na$_2$O-80SiO$_2$ glass at 300 K. Sketches of the characteristic regions embedded in the structure and the corresponding persistence diagram of these loops are also shown in Fig. 1. The defined regions have different attributes owing to the different structural characteristics. C_L is where loops are born early and have a long lifetime (close to the vertical axis). These are chains, for which the connected atoms are in spatial proximity (e.g., covalent Si-O structures or space-filling Na atoms at higher Na$_2$O contents). Furthermore, the large lifetime of C_L indicates a large separation for the most distant atoms of the loops, typically found on opposite side of the loops. C_4 is where loops have varying birth and death times. These are primarily ascribed to packed structures of oxygen and sodium atoms. C_3 and C_R have varying birth times and short lifetimes (i.e. they are positioned close the diagonal). These are typically similar in composition to C_4 but often more separated in space as seen by the later birth time. Last, C_T has both low birth time and low death time (close to the origin), emerging from the intra-tetrahedral O-Si-O structure.

The topological implication of the sodium incorporation in the silica network is shown in Fig. 3, where the characteristic regions are observed to gradually change upon Na$_2$O addition (the remaining persistence diagrams are shown in fig. S6). The number of loops in the low–birth time regions of C_L, C_3, and C_4 increases as a result of Na$_2$O addition (Fig. 3). We analyzed these loops and found them to be primarily oxygen (~75% in average, decreasing with increasing sodium content) with smaller amounts of sodium and silicon. The decreasing birth time of these loops is due to the denser packing
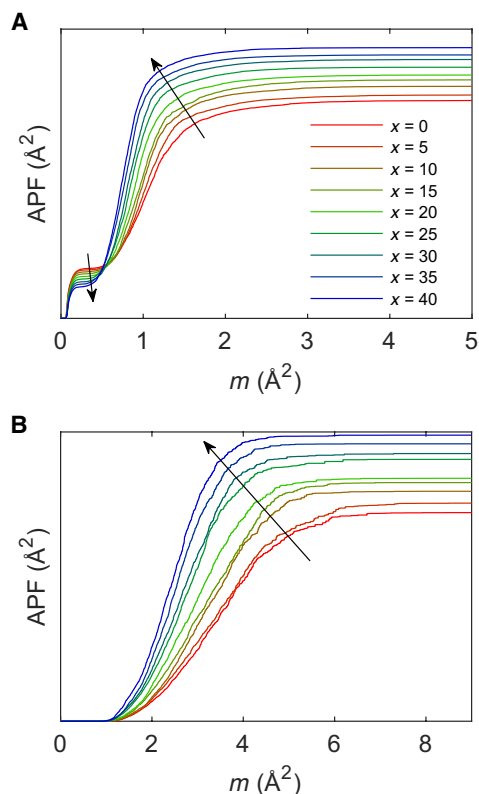
of oxygen atoms around sodium cations in the structure, providing shorter separation distances, in agreement with the partial RDFs (Fig. 2, A to C, and fig. S5, A to C). Such decrease of interatomic separation will decrease both birth and death times. The lifetime of C_L loops is also found to decrease with increasing Na$_2$O content, as closed Si-O loops are expected to occur less frequently with increasing depolymerization of silicon-oxygen structures (fig. S2) and void filling by Na$^+$ ions. Generally, the growth of the C_L, C_3, and C_4 regions with increasing Na$_2$O content is expected, given the increasing number of new structural features of the modified glasses, yet it demonstrates how persistent homology may be used to discover new MRO features in amorphous materials.

Quantitative comparison of persistence diagrams

While persistent homology is a powerful tool for studying amorphous structures, comparisons between persistence diagrams (as in Fig. 3) have previously been qualitative (21). To resolve this, we here compute the APF function (Eq. 1) for both loops (Fig. 4A) and voids (Fig. 4B) of the studied glasses. Figure 4A essentially compiles the information of Fig. 3 and fig. S6 into a single graph, by accumulating lifetimes as a function of the mean age, $m = (b + d)/2$, enabling quantitative comparisons. The shape of the APF curves is qualitatively similar for all the compositions, with an initial inflection point at $m \sim 0.25$ Å^2 and a following critical point where the APF intensity decreases with increasing Na$_2$O content. As m increases further, the

Fig. 3. Persistence diagrams of loops of selected sodium silicate glasses. Persistence diagrams for structures at 300 K with $x = \{0,10,20,40\}$ in the xNa$_2$O-(100 − x)SiO$_2$ system.

Fig. 4. Quantitative analysis of persistence diagrams using the APF. The APF is presented for (**A**) loops and (**B**) voids as calculated from Eq. 1 for glasses $x = \{0,5,10,15,20,25,30,35,40\}$ in the xNa$_2$O-(100 − x)SiO$_2$ system showing noticeable differences in MRO of the studied glasses. Arrows indicate tendency with increasing Na$_2$O content.

APF rises steeply to a maximum value, which is found at smaller m values for increasing Na_2O content. We furthermore note that this maximum value increases with increasing Na_2O content. The first inflection point at $m \sim 0.25$ $Å^2$ may be explained by differences in C_T related to the reduced number of Si atoms in the simulation, as this region primarily originates from the intra-tetrahedral three-atom O-Si-O structure (see Fig. 1 for a sketch of this structure). In contrast, we ascribe the second steep increase in the APF with m to the growth of C_L, C_3, and C_4. These regions primarily consist of loops with all three atom types, yet with a majority of oxygen. Hence, the steep increase is due to an increased number of loops and not an increased lifetime of loops. Another important observation is the increase of the APF past the first inflection point, which occurs earlier with increasing Na_2O content due to a decreased mean age of the loops. This corresponds to the data points (loops) moving closer to the origin of the plot in Fig. 3 with increasing Na_2O content. These results for C_3, C_4, and C_R may be understood based on the effect of adding sodium, which fills structural voids of silicon and oxygen packing, hence creating new loops, consisting almost exclusively of O and Na. In turn, these are subloops of C_L but with shorter mean separation across the loop compared to pure O subloops and thus shorter death time and mean age. A similar decrease of mean age is also found for C_L as a result of the incorporation of sodium into the loops. This is in good agreement with the decreasing number of fully connected silicon tetrahedra (Q^4 units, as seen in fig. S2) as well with the changes of the partial RDFs of Fig. 2 (A to C) showing noticeable structural compaction.

Clear differences among the glasses are also noted in the APFs of the network voids (Fig. 4B). The corresponding persistence diagrams of voids, which have been used to construct Fig. 4B, are presented in fig. S7 and generally show slightly decreasing mean age and lifetime with increasing Na_2O content. Structurally, this corresponds to smaller voids, in agreement with the increasing packing factor as both found experimentally (34) and in the present simulations (fig. S3).

While the ability of the APF to compare persistence diagrams is inherently useful, its additive nature also enables the contribution from each characteristic region to be deconvoluted. We provide an example of such analysis in fig. S8, finding C_L and C_4 to only contribute to the second inflection of the APF, while C_3 and C_R contribute to both inflections. Overall, we find that the combination of the persistence diagram and APF is a valuable tool for probing MRO and atomic packing in complex amorphous systems. That is, while visual, qualitative inspection of the persistence diagrams allows for some categorization of the different characteristic regions, the comparison of APFs offers a robust method for quantitatively analyzing differences in the contributions of the characteristic regions to the persistence diagram.

## Deconvoluting the FSDP by persistent homology

The obtained persistence diagrams also hold quantitative information about loop/ring sizes (as measured by the death time) that can be related with experimental observables, such as the FSDP. To do so, we here use Eq. 2 to calculate the $S_{PH}(Q)$ function of the characteristic regions of the persistence diagram for a $20Na_2O$-$80SiO_2$ glass and compare the results with $S(Q)$ (Fig. 5). The latter is computed by the Fourier transform of the RDF using a Lorch-type window function to reduce ripples at low Q values (40). Note that we exclude contributions of C_T (i.e., contributions of the intra-tetrahedral O-Si-O loops) from all $S_{PH}(Q)$ function analyses to restrict the analyses
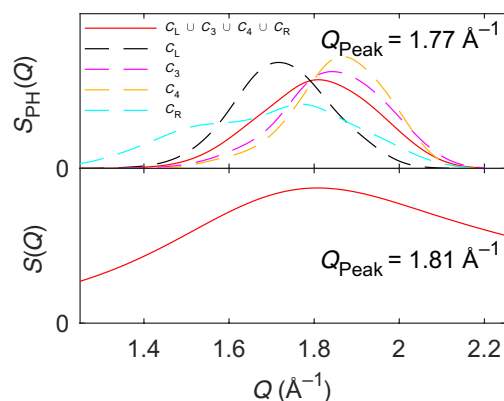
Fig. 5. Deconvolution of the FSDP. Comparison between the $S_{PH}(Q)$ function calculated from Eq. 2 for all characteristic regions as well as their union and the structure factor [$S(Q)$], both shown for the $20Na_2O$-$80SiO_2$ glass. Note that we exclude C_T from this analysis.

to MRO features and that we use equal scattering lengths for all atomic pairs when comparing $S(Q)$ to $S_{PH}(Q)$. Overall, we observe substantial differences in peak position among the different characteristic regions (highest Q value for C_4 and lowest for C_R). A comparison of the $S_{PH}(Q)$ function of the union $A_k = C_L \cup C_3 \cup C_4 \cup C_R$ with the computed $S(Q)$ (see Fig. 5) reveals excellent agreement of peak positions (1.77 and 1.81 $Å^{-1}$, respectively), providing evidence for the applicability of this approach to understand the origin of the FSDP.

To further understand the contributions to the FSDP, we calculate the $S_{PH}(Q)$ function for the characteristic regions C_L, C_3, C_4, and C_R as well as the union of these for all the studied glasses in the xNa_2O-$(100 - x)SiO_2$ series. The $S_{PH}(Q)$ functions of C_L, C_3, and $C_L \cup C_3 \cup C_4 \cup C_R$ are shown in Fig. 6 (A to C), while the remaining ones are shown in fig. S9. The peak positions of the $S_{PH}(Q)$ function for C_L (Fig. 6A), C_3 (Fig. 6B), and $C_L \cup C_3 \cup C_4 \cup C_R$ (Fig. 6C) are all found to increase with increasing Na_2O content. Comparing these $S_{PH}(Q)$ functions with the computed $S(Q)$ in Fig. 6D, we observe the same compositional dependence (increasing Q position with increasing sodium content), thus confirming the applicability of the proposed method to understand the FSDP. We directly compare the maximum position of $S(Q)$ to the weighted average of the various $S_{PH}(Q)$ functions in fig. S10, finding the best agreement for C_4 and the union $C_L \cup C_3 \cup C_4 \cup C_R$ (fig. S10, C and E, respectively). This highlights the need for considering all characteristic regions in the description of the FSDP. We note some intensity differences between $S_{PH}(Q)$ functions and $S(Q)$ and ascribe these to how $S(Q)$, unlike $S_{PH}(Q)$, incorporates scattering interference contributions. This is supported by the work of Du and Corrales (12, 13), which found the FSDP intensity of three alkali (Li, Na, and K) silicates to mainly vary because of the difference in their neutron scattering lengths and not because of differences in their MRO. That is, the FSDP intensity of lithium silicate (lithium has a negative neutron scattering length) varies inconsiderably with increasing Li_2O content, while the FSDP intensities of sodium and potassium (both have positive neutron scattering lengths of similar magnitude) silicates show a substantial and similar decrease upon modifier addition. Similarly, the position of the FSDP has been found to only vary unnotably when changing the alkali ion (13), highlighting the pronounced effect of scattering interference in the obtained FSDP intensity from Fourier transform of the RDF.
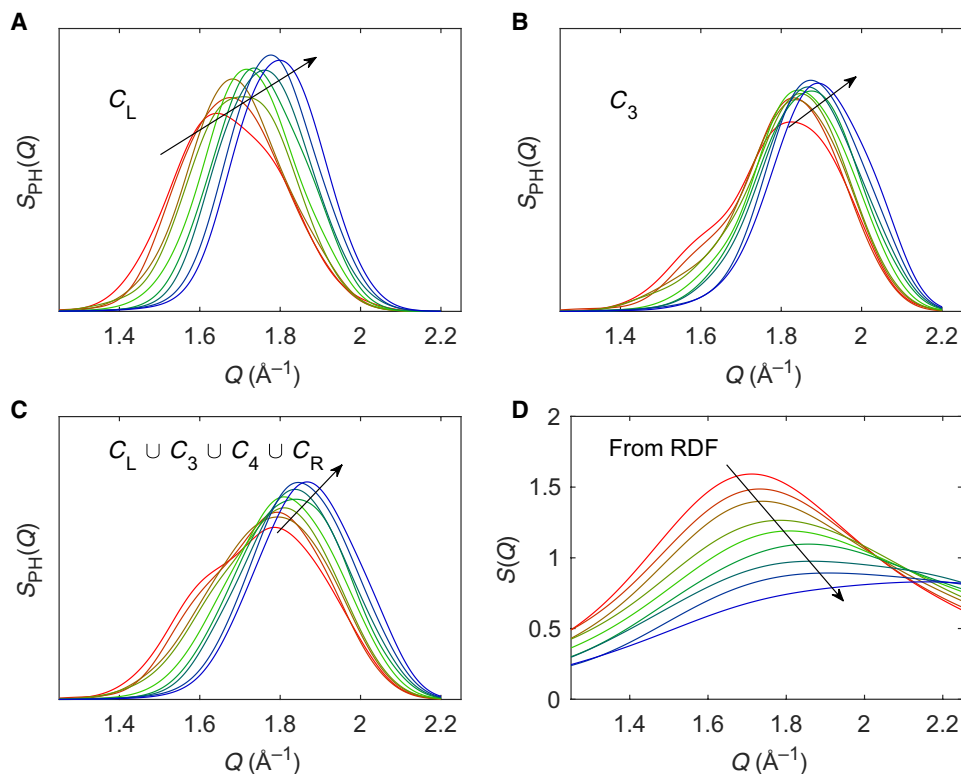
Fig. 6. Composition-dependent changes in the contributions of various characteristic regions on the FSDP. A comparison between the $S_{PH}(Q)$ function as computed from different characteristic regions; (**A**) C_L, (**B**) C_3, (**C**) the union $C_L \cup C_3 \cup C_4 \cup C_R$ of the persistence diagrams, and (**D**) the FSDP of the structure factor [$S(Q)$] obtained from the RDFs of structures from molecular dynamics simulations. The color coding is similar to Fig. 4, and arrows indicate tendency with increasing Na_2O content.

On the basis of the above analysis, we conclude that the compositional dependence of the FSDP in sodium silicate glasses is related to all of the studied characteristic regions but with more pronounced contributions from large (>3 atoms) loops. Given the major abundance of oxygen in all these regions and the large radii of O and Na compared to Si, we ascribe the FSDP in the present case to originate from structural features of oxygen/sodium packing. This result is consistent with the OO partial structure factor (Fig. 2F), which is the only pair correlation to feature the Q correlation seen in Fig. 6D. The structural origin of these observations is likely ascribed to the packing of Na^+ in the voids of the silicate backbone structure. Within the structure, the cationic sodium provides additional bonding between otherwise repulsive parts of the Si-O network. This reduces the distance between oxygen atoms due to their shared coordination to Na^+ (see Fig. 1 for a schematic illustration of such packing). This effect becomes enhanced at higher sodium content. Similarly to the oxygen-oxygen correlation, the NaO partial structure factor (Fig. 2D) also shows a shift and a decrease in intensity with increasing Na_2O content. Furthermore, these considerations are in good agreement with the partial OO RDF (Fig. 2C), showing a clear growth of the second peak at ~3.5 Å and a related decrease in intensity of the third peak at ~5 Å upon increasing Na_2O content. Such compaction of the structure with increasing sodium content is also seen from the increase in intensity of the first peak in the NaO correlation (Fig. 2A) and the decrease in correlation length of the NaNa correlation (fig. S5A). We note that these results agree with the decrease in the mean age (i.e., atoms are less separated on average) of the second inflection point of the APF (Fig. 4A), the steeper slope during

the APF increase (Fig. 4A), and the decreasing void size (Fig. 4B). Last, the reduced separation is in agreement with the increasing Q values of the FSDP in inverse space upon increasing Na_2O content (Fig. 6D).

DISCUSSION

We first compare the present persistent homology approach for analyzing the FSDP with other models. The model of Shi *et al.* (*3*) relates the silica ring size distributions from silicate crystals to the glassy FSDP. In comparison, the present approach has the advantage of not having any crystallography-based assumptions regarding the diameter of the rings. Similarly, the present approach allows for more direct structural deconvolution than the questionable prepeak in the concentration-concentration structure factor (*17*, *18*) considered in Elliot's void model (*6*). Last, while the recent model of Shi and Tanaka (*4*) offers new insights into the contribution of the tetrahedral unit to the FSDP, the present approach is more widely applicable also to nontetrahedral structures. Several other approaches for explaining the FSDP exist, mainly based on describing the FSDP through the size and distribution of nanovoids in the studied materials (*2*, *41*). Such models generally rely on first defining the voids and then their size, typically requiring a number of assumptions on, e.g., chemical bonding. The persistent homology approach greatly reduces the number of such assumptions (and thus the potential bias) to the atomic radii and inherently includes all atom types in the computation, with a description of the loop/ring size through the death time. As such, the advantage of our developed persistent homology-based

methodology lies in its applicability to any amorphous system, as it only requires atomic trajectory and radii information, without requirements of prior knowledge of chemical bonding.

In comparison to the previous work on persistent homology, this work considerably expands the number of amorphous structures studied by the S_{PH} function (from 3 to 12) (21, 23) and provides the first rigorous test of the methodology for a systematic composition variation in a model glass series. As such, we have generalized the use of persistent homology to deconvolute the contribution of various MRO features to the FSDP in oxide glasses. By introducing self-consistent rigorous geometrical definitions of the structural units compared to previous works, we decode the persistence diagrams into structural features with simple geometrical origin. In addition, by implementing the so-called APF as a quantitative analysis tool for amorphous structures, we are able to decipher how the structural features identified by persistent homology evolve as a function of, e.g., time, stress, or composition. Consequently, we believe that this approach may lead to the settlement of the long-lasting debate regarding the origin of the FSDP in glassy systems. More generally, it can be applied to analyze and categorize molecular dynamics data and reveal hidden MRO in various amorphous materials.

METHODS
Sodium silicate glass structures (3000 atoms) were simulated using the potential of Teter (32, 42), following the quenching procedure in (43). Comparisons of experimental and simulated neutron $S(Q)$ and Q^n speciation are shown in figs. S1 and S2, respectively. We note a slight discrepancy in the FSDP between experiments and molecular dynamics. However, this has no influence on conclusions of this work, which only deals with comparisons of the FSDP from the RDF and persistent homology, both based on the same molecular dynamics simulations. Furthermore, the compositional dependence of density and packing factor, as well as the Si-O ring size distribution, is shown in figs. S3 and S4. We generally find good agreement between experiments and simulations. We computed the partial radial distributions and structure factors functions as averages over 100 ps of simulation time. This was done to obtain partial NaNa RDFs of meaningful quality for the structures of lowest sodium concentration. We found negligible differences in the partial structure factors and the total FSDP position when comparing static and time-averaged structure factors.

Persistent homology analyses were conducted using the libraries Diode (44) and Dionysus 2 (45), which incorporate computation based on periodic boundary conditions. We have computed the presented persistent homology functions using both energy-minimized (i.e., inherent configuration) and unminimized structures of all glass compositions, showing negligible differences. All presented data are thus for unminimized structures. In the treatment of the atomic radii (r), we follow the approach in (21) to obtain $r_O = 1.289$ Å, $r_{Na} = 1.070$ Å, and $r_{Si} = 0.326$ Å.

SUPPLEMENTARY MATERIALS

REFERENCES AND NOTES
1. S. R. Elliott, Medium-range structural order in covalent amorphous solids. *Nature* **354**, 445–452 (1991).
2. J. M. Zaug, A. K. Soper, S. M. Clark, Pressure-dependent structures of amorphous red phosphorus and the origin of the first sharp diffraction peaks. *Nat. Mater.* **7**, 890–899 (2008).
3. Y. Shi, J. Neuefeind, D. Ma, K. Page, L. A. Lamberson, N. J. Smith, A. Tandia, A. P. Song, Ring size distribution in silicate glasses revealed by neutron scattering first sharp diffraction peak analysis. *J. Non Cryst. Solids* **516**, 71–81 (2019).
4. R. Shi, H. Tanaka, Distinct signature of local tetrahedral ordering in the scattering function of covalent liquids and glasses. *Sci. Adv.* **5**, eaav3194 (2019).
5. Y. Q. Cheng, E. Ma, Atomic-level structure and structure-property relationship in metallic glasses. *Prog. Mater. Sci.* **56**, 379–473 (2011).
6. S. R. Elliott, Origin of the first sharp diffraction peak in the structure factor of covalent glasses. *Phys. Rev. Lett.* **67**, 711–714 (1991).
7. C. Crupi, G. Carini, M. González, G. D'Angelo, Origin of the first sharp diffraction peak in glasses. *Phys. Rev. B* **92**, 134206 (2015).
8. L. E. Busse, S. R. Nagel, Temperature dependence of the structure factor of As$_2$Se$_3$ glass up to the glass transition. *Phys. Rev. Lett.* **47**, 1848–1851 (1981).
9. C. Lin, L. E. Busse, S. R. Nagel, Temperature dependence of the structure factor of GeS$_2$ glass. *Phys. Rev. B* **29**, 5060–5062 (1984).
10. K. Tanaka, Pressure dependence of the first sharp diffraction peak in chalcogenide and oxide glasses. *Philos. Mag. Lett.* **57**, 183–187 (1988).
11. Y. Inamura, Y. Katayama, W. Utsumi, K.-i. Funakoshi, Transformations in the intermediate-range structure of SiO$_2$ glass under high pressure and temperature. *Phys. Rev. Lett.* **93**, 015501 (2004).
12. J. Du, L. R. Corrales, First sharp diffraction peak in silicate glasses: Structure and scattering length dependence. *Phys. Rev. B* **72**, 092201 (2005).
13. J. Du, L. R. Corrales, Compositional dependence of the first sharp diffraction peaks in alkali silicate glasses: A molecular dynamics study. *J. Non Cryst. Solids* **352**, 3255–3269 (2006).
14. P. H. Gaskell, D. J. Wallis, Medium-range order in silica, the canonical network glass. *Phys. Rev. Lett.* **76**, 66–69 (1996).
15. A. C. Wright, R. N. Sinclair, A. J. Leadbetter, Effect of preparation method on the structure of amorphous solids in the system As-S. *J. Non Cryst. Solids* **71**, 295–302 (1985).
16. L. Červinka, Medium-range order in amorphous materials. *J. Non Cryst. Solids* **106**, 291–300 (1988).
17. C. Massobrio, A. Pasquarello, Origin of the first sharp diffraction peak in the structure factor of disordered network-forming systems: Layers or voids? *J. Chem. Phys.* **114**, 7976–7979 (2001).
18. Q. Mei, C. J. Benmore, S. Sen, R. Sharma, J. L. Yarger, Intermediate range order in vitreous silica from a partial structure factor analysis. *Phys. Rev. B* **78**, 144204 (2008).
19. M. Micoulaut, Diffraction patterns of amorphous materials as a series expansion of neighbor distribution functions. *J. Phys. Condens. Matter* **31**, 285402 (2019).
20. M. Murakami, S. Kohara, N. Kitamura, J. Akola, H. Inoue, A. Hirata, Y. Hiraoka, Y. Onodera, I. Obayashi, J. Kalikka, N. Hirao, T. Musso, A. S. Foster, Y. Idemoto, O. Sakata, Y. Ohishi, Ultrahigh-pressure form of SiO$_2$ glass with dense pyrite-type crystalline homology. *Phys. Rev. B* **99**, 045153 (2019).
21. Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escolar, K. Matsue, Y. Nishiura, Hierarchical structures of amorphous solids characterized by persistent homology. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7035–7040 (2016).
22. S. Hosokawa, J.-F. Bérar, N. Boudet, W.-C. Pilgrim, L. Pusztai, S. Hiroi, K. Maruyama, S. Kohara, H. Kato, H. E. Fischer, A. Zeidler, Partial structure investigation of the traditional bulk metallic glass Pd$_{40}$Ni$_{40}$P$_{20}$. *Phys. Rev. B* **100**, 054204 (2019).
23. S. Hong, D. Kim, Medium-range order in amorphous ices revealed by persistent homology. *J. Phys. Condens. Matter* **31**, 455403 (2019).
24. M. R. Mcguirl, A. Volkening, B. Sandstede, Topological data analysis of zebrafish patterns. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5113–5124 (2020).
25. G. Carlsson, Topology and data. *Bull. Am. Math. Soc.* **46**, 255–308 (2009).
26. C. A. N. Biscio, J. Møller, The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications. *J. Comput. Graph. Stat.* **28**, 671–681 (2019).
27. H. Edelsbrunner, J. Harer, *Computational Topology: An Introduction* (American Mathematical Society, 2009).
28. D. A. Keen, A comparison of various commonly used correlation functions for describing total scattering. *J. Appl. Cryst.* **34**, 172–177 (2001).
29. T. E. Faber, J. M. Ziman, A theory of the electrical properties of liquid metals. *Philos. Mag.* **11**, 153–173 (1965).
30. S. Le Roux, P. Jund, Ring statistics analysis of topological networks: New approach and application to amorphous GeS$_2$ and SiO$_2$ systems. *Comput. Mater. Sci.* **49**, 70–83 (2010).
31. N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, H. A. Harrington, A roadmap for the computation of persistent homology. *EPJ Data Sci.* **6**, 17 (2017).
32. J. Du, in *Molecular Dynamics Simulations of Disordered Materials* (Springer, 2015), pp. 157–180.

33. M. Fábián, P. Jóvári, E. Sváb, G. Mészáros, T. Proffen, E. Veress, Network structure of 0.7SiO$_2$-0.3Na$_2$O glass from neutron and x-ray diffraction and RMC modelling. *J. Phys. Condens. Matter* **19**, 335209 (2007).

34. S. Giri, C. Gaebler, J. Helmus, M. Affatigato, S. Feller, M. Kodama, A general study of packing in oxide glass systems containing alkali. *J. Non Cryst. Solids* **347**, 87–92 (2004).

35. M. Misawa, D. L. Price, K. Suzuki, The short-range structure of alkali disilicate glasses by pulsed neutron total scattering. *J. Non Cryst. Solids* **37**, 85–97 (1980).

36. V. F. Sears, Neutron scattering lengths and cross sections. *Neutron News* **3**, 26–37 (1992).

37. W.-A. Buckermann, W. Müller-Warmuth, G. H. Frischat, A further ^{29}Si MAS NMR study on binary alkali silicate glasses. *Glas. Berichte.* **65**, 18–21 (1992).

38. A. C. Wright, The comparison of molecular dynamics simulations with diffraction experiments. *J. Non Cryst. Solids* **159**, 264–268 (1993).

39. R. D. Shannon, Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr.* **32**, 751–767 (1976).

40. A. C. Wright, Neutron and X-ray amorphography. *J. Non Cryst. Solids* **106**, 1–16 (1988).

41. M. A. Popescu, Hole structure of computer models of non-crystalline materials. *J. Non Cryst. Solids* **35–36**, 549–554 (1980).

42. S. Plimpton, Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).

43. X. Li, W. Song, K. Yang, N. M. A. Krishnan, B. Wang, M. M. Smedskjaer, J. C. Mauro, G. Sant, M. Balonis, M. Bauchy, Cooling rate effects in sodium silicate glasses: Bridging the gap between molecular dynamics simulations and experiments. *J. Chem. Phys.* **147**, 074501 (2017).

44. Diode—Persistent homology software; https://github.com/mrzv/diode.

45. Dionysus2—Persistent Homology software; https://mrzv.org/software/dionysus2/.