


## Statistical topology of bond networks with applications to silica

B. Schweinhart \**Department of Mathematics, Ohio State University, Columbus, Ohio 43210, USA*D. Rodney<sup>†</sup>*Institut Lumière Matière, University of Lyon, Villeurbanne, Lyon F-69622, France*J. K. Mason ‡*Department of Materials Science and Engineering, University of California, Davis, California 95616, USA*

(Received 27 October 2019; accepted 9 March 2020; published 19 May 2020)

Whereas knowledge of a crystalline material's unit cell is fundamental to understanding the material's properties and behavior, there are no obvious analogs to unit cells for disordered materials despite the frequent existence of considerable medium-range order. This article views a material's structure as a collection of local atomic environments that are sampled from some underlying probability distribution of such environments, with the advantage of offering a unified description of both ordered and disordered materials. Crystalline materials can then be regarded as special cases where the underlying probability distribution is highly concentrated around the traditional unit cell. The  $H_1$  barcode is proposed as a descriptor of local atomic environments suitable for disordered bond networks and is applied with three other descriptors to molecular dynamics simulations of silica glasses. Each descriptor reliably distinguishes the structure of glasses produced at different cooling rates, with the  $H_1$  barcode and coordination profile providing the best separation. The approach is generally applicable to any system that can be represented as a sparse graph.

DOI: [10.1103/PhysRevE.101.052312](https://doi.org/10.1103/PhysRevE.101.052312)

### I. INTRODUCTION

A bond network is one in which atoms connected by covalent bonds form a network that extends throughout the material. The nature of covalent bonding is such that every atom of a given species generally forms the same number of covalent bonds (referred to below as the *valence*), though Fig. 1 shows that this does not significantly restrict the overall network connectivity. The purpose of this article is to characterize this connectivity and to enable quantitative comparison of the connectivity of different bond networks. Since the connectivity is entirely defined by knowledge of the existence of atoms and the bonds connecting them, the resulting analysis does not depend on variations in bond energies or the precise geometries of local atomic environments. That is, a bond network is considered as a graph (in the sense used in discrete mathematics) where the atoms constitute the vertices (possibly labeled by atomic species) and the covalent bonds the edges. Some geometric information is implicitly retained, as the bond lengths and bond angles constrain the allowable network topologies and vice versa [1,2]. It is partly for this reason that other purely topological quantities such as coordination number and ring statistics are commonly used to quantify the structure of bond networks and why we propose

that our methodology will provide an informative language to describe these systems.

While this article specifically considers local atomic environments in silica glasses, the approach is generally applicable to any system that can be represented as a sparse graph (one with low-valence vertices). A bond network for, e.g., a metallic glass [3], could be defined by constructing the Voronoi tessellation and identifying every face of a Voronoi polyhedron with a bond. Similarly, tile edges in a Penrose tiling [4], films in a two-dimensional soap froth, and grain boundaries in a polycrystalline thin film [5] could all be considered as bonds connecting corners or junction points. It is true that considering only the graph structure of these systems discards the geometric information, but this can be an advantage if the purpose is to identify exceptional local environments. Such environments generally have an associated topological signature, and considering only the local graph structure can simplify the analysis significantly. Moreover, retaining excessive local information can actually frustrate the users ability to distinguish disordered structures for reasons explored in Sec. III B.

Let the distance between two atoms be defined as the number of edges along the shortest path between the corresponding vertices. A local atomic environment of radius  $r$  centered at an atom  $v$  is then defined as the subgraph consisting of all atoms within distance  $r$  of  $v$  and all covalent bonds between atoms in this set, as in Fig. 2. The atom  $v$ , with a distinguished position at the center of the local atomic environment, is called the *root*. A perfect crystalline solid, by definition,

\*schweinhart.2@osu.edu

†david.rodney@univ-lyon1.fr

‡jkmason@ucdavis.edu

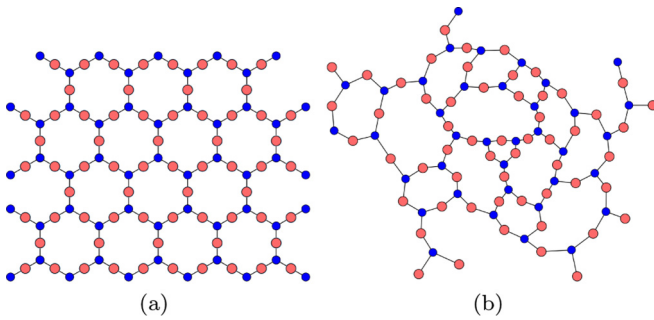


FIG. 1. Ordered (a) and disordered (b) bond networks containing 2-valent red (light gray) atoms and 3-valent blue (dark gray) atoms.

contains a small number of topological types of local atomic environments; considering the polymorphs of silica ( $\text{SiO}_2$ ) as examples, there is just one environment in cristobalite and two in coesite if the roots are restricted to silicon atoms. By contrast, a disordered solid contains many topological types, although these do not all necessarily have the same probability of occurrence. This suggests that every material with a given chemical composition and processing history has a characteristic probability distribution of local atomic environments, with crystalline solids as special cases where the probability distribution is highly concentrated. Knowledge of this probability distribution would then subsume that of the unit cell and apply to ordered and disordered bond networks alike. This approach has the further advantage of being sensitive to both continuous and discontinuous variations in the rates of occurrence of local atomic environments and could therefore be used for, e.g., phase detection, identification of crystal nuclei, or verification that computationally generated structures correspond to experimental ones.

A central concern that has yet to be addressed is precisely which local atomic environments should be considered equivalent and which distinct. We propose that there is no single notion of equivalence that is preferable in all situations. Four possibilities are discussed below in the context of molecular dynamics simulations of silica glasses produced at different cooling rates. For this application, the equivalence relation that allows the atomic networks to be most easily distinguished on the basis of local structural differences is preferred. If instead the relative nucleation rates of competing crystalline phases was the subject of study, then perhaps the one that could most easily distinguish the characteristic

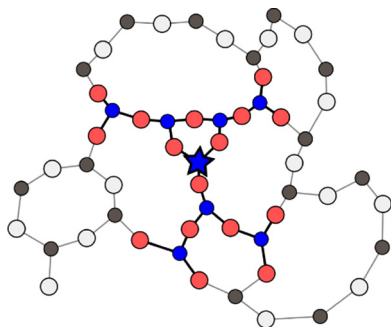


FIG. 2. A local atomic environment of radius 5 inside a larger bond network. The root atom is marked by a star.

probability distributions of different crystalline phases would be more preferable.

In previous work [6,7], two local environments were considered to be equivalent up to graph isomorphism, effectively employing all of the available topological information to compare local atomic environments. While theoretically satisfying, this can cause the number of equivalence classes (distinct types of local environments) to grow very quickly with the radius of the neighborhood. For example, using this equivalence relation gives  $9 \times 10^4$  distinct equivalence classes in a sample of  $10^5$  environments in silica glass at a radius sufficient to differentiate the crystalline forms of  $\text{SiO}_2$ . This is not a useful summary of the local structure; the true probability distribution of local atomic environments would require an infeasibly large sample population to estimate and even if known would consider nearly every environment as unique.

This proliferation of equivalence classes is addressed by considering three further equivalence relations which are coarser (encode less information) than graph isomorphism and effectively assign a summary of topological properties to a local atomic environment. Two are based on classical descriptors, namely the coordination profile which records the coordination numbers of atoms at each distance from the root, and the primitive ring profile, which records the number and length of primitive rings containing the root. The third equivalence relation is the  $H_1$  barcode, a novel ring-based descriptor that is defined in terms of the first homology group. The  $H_1$  barcode is better at distinguishing local atomic environments in crystalline and glassy silica than the primitive ring profile, particularly when every atom in the network is required to have the standard valence.

There is an extensive literature on local structural descriptors for bond networks (the introduction of Ref. [7] gives a brief survey). Specifically for covalent glasses, our approach is most similar to the local clusters of Hobbs *et al.* [8]. The local cluster of an atom is the union of all primitive rings (defined below) containing that atom, and the equivalence relation on local clusters is graph isomorphism. This approach is, however, found to suffer from the same proliferation of equivalence classes described above for the graph isomorphism equivalence relation. Ring statistics are perhaps the most commonly used local structural descriptor for silica networks [9], with a survey of several different classes of distinguished rings given in Ref. [10]. We study one descriptor based on ring statistics, the primitive ring profile. The primitive ring data are used to classify local environments into different equivalence classes rather than to compute global ring statistics though, providing information about the adjacency of rings and allowing for different methods of analysis. Recently, persistent homology was applied to the study silica networks [11]; while the  $H_1$  barcode introduced here can be defined in terms of persistent homology, it gives qualitatively different information than the methodology of Ref. [11].

Previous studies have identified various structural properties of silica glasses that depend on, e.g., the cooling rate from the liquid [12–16]. Our methodology is applied to simulations of silica glasses quenched from the liquid at three different rates. Overall, glasses that form with faster cooling rates are found to exhibit more disorder, have a higher frequency of rings with six or eight atoms (three or four silicon atoms), and have more coordination defects.

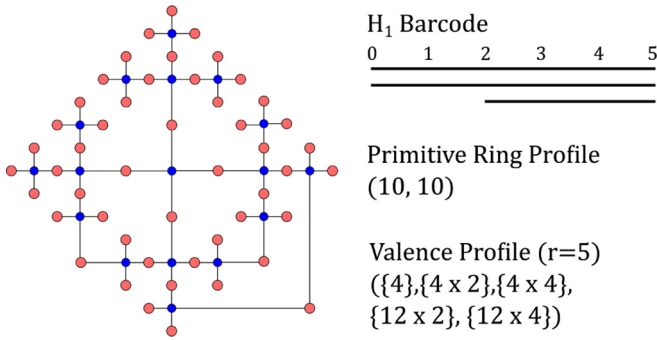


FIG. 3. An example local atomic environment and its  $H_1$  barcode, primitive ring profile, and coordination profile.

## II. CLASSIFICATION OF LOCAL ATOMIC ENVIRONMENTS

Let  $V$  be the set of atoms comprising a bond network. A structural descriptor of a bond network is defined as a function  $X$  that assigns to the atom  $v$  a summary  $X(v)$  of the properties of the local atomic environment around  $v$ . This induces an equivalence relation on atoms  $v \in V$  in the sense that atoms  $v_1$  and  $v_2$  are considered to be equivalent if  $X(v_1) = X(v_2)$ . Given two descriptors  $X$  and  $Y$ ,  $X$  is said to be coarser than  $Y$  if equivalence under  $Y$  implies equivalence under  $X$  for all pairs of atoms (i.e., the information provided by  $Y$  contains that provided by  $X$ ). Each of the descriptors considered here depends on the radius  $r$  of the local environment, and each has the property that  $X_r$  is coarser than  $X_{r+1}$  for all  $r > 0$  (i.e., higher values of the radius provide more information).

Let the *equivalence class*  $x$  denote the set of all atoms  $v$  such that  $X(v) = x$ . Since the underlying probability distribution on equivalence classes is not usually known for a given descriptor  $X$  and atom set  $V$ , our approach instead uses empirical probability distributions. The empirical probability distribution of  $X$  on  $V$  assigns to every equivalence class  $x$  the probability that  $X(v) = x$  for a randomly chosen atom  $v \in V$ , or

$$P_X(x) = |\{v \in V : X(v) = x\}|/|V|.$$

A descriptor  $X$  and two bond networks with atom sets  $V_1$  and  $V_2$  then result in two discrete probability distributions  $P_X(x)$  and  $Q_X(x)$ , and a measure of the similarity of these discrete probability distributions could be used to define the similarity of the bond networks (e.g., a metric on discrete probability distributions induces a pseudometric on bond networks). While the symmetrized Kullback-Leibler (KL) divergence [17] is used for this purpose below, other possibilities include the standard  $L_1$  or  $L_2$  metrics or a Wasserstein distance that incorporates information about the geometric similarity of equivalence classes [7].

Three of the four descriptors below encode intuitive topological information about a local atomic environment, although the technical descriptions of the descriptors can be deceptively involved. It is useful to frequently consult examples of local atomic environments and the corresponding descriptors while reading the descriptions below to develop the above-mentioned intuition. One example is provided in Fig. 3, and many more of relevance to silica glasses are given in Appendix D.

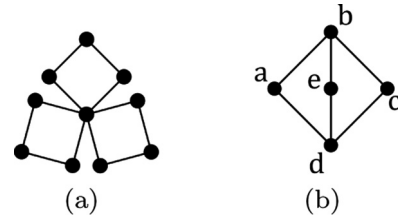


FIG. 4. Two configurations with three primitive rings of length four. Panel (a) has three algebraically independent rings and panel (b) has only two.

### A. Graph isomorphism

The most detailed of the descriptors uses *graph isomorphism* to construct equivalence classes of local atomic environments. This requires that two local atomic environments be considered equivalent if there is an invertible mapping of atoms from one to the other that preserves atomic types and bonds. The resulting equivalence class specifies all topological information about the local atomic environment. From the standpoint of ring statistics, graph isomorphism controls the number and length of rings as well as their adjacency and distance from the root. The methodology described in Ref. [7] was used to calculate the corresponding descriptor; that is, a local environment was represented by an adjacency matrix written in canonical form using the software package Nauty [18].

### B. The coordination profile

The *coordination profile* is the simplest descriptor considered here. At radius  $r$  this is a vector of  $r$  unordered lists where the  $i$ th list gives the coordinations (valences) of all atoms at distance  $i$  from the root. An example of a local atomic environment and its coordination profile is shown in Fig. 3; the root has valence four, each of the atoms in the first neighbor shell has valence two, and the valences alternate with neighbor shell number. Observe that the coordination profile at radius  $r$  implicitly includes information about the bonds between atoms in the  $r$ th and  $(r+1)$ th shells; this means that the coordination profile at radius  $r$  provides a coarser classification than graph isomorphism at radius  $r+1$ , but not at radius  $r$ .

For a perfectly coordinated silica network—one where every oxygen is adjacent to two silicons and every silicon is adjacent to four oxygens—the coordination profile contains equivalent information to the number of atoms in each neighbor shell around the root. This is known as the *shell count* of the local atomic environment. For example, the radius 5 shell count of the configuration in Fig. 3 is (1, 4, 4, 12, 12, 33).

### C. The primitive ring profile

A *primitive ring* [19,20] is defined as a ring of bonded atoms  $(v_1, v_2, \dots, v_k, v_1)$  such that the shortest path connecting any pair of atoms in the ring is contained within the ring. That is, the shortest path between  $v_i$  and  $v_j$  for all  $1 \leq i < j \leq k$  is either  $v_i, v_{i+1}, \dots, v_j$  or  $v_j, \dots, v_k, v_1, v_2, \dots, v_i$ . For example, the configurations in Fig. 4 both have three primitive rings. Note that if  $v$  is an atom in a bond network, then a primitive ring containing  $v$  is primitive with respect to the local atomic environment rooted at  $v$  if and only if it is

primitive in the entire network as well; this is not necessarily true if a ring does not contain  $v$ . The *primitive ring profile* of radius  $r$  of an atom  $v$  gives the lengths of all primitive rings which contain  $v$  and at most  $2r$  total atoms, as in the example in Fig. 3. Unlike the coordination profile, the primitive ring profile stabilizes at a relatively small value of the radius  $r$  (there are often alternate paths connecting distant parts of large rings). The algorithm of Yuan and Cormack [21] is used to compute primitive rings.

The ring lengths given here are *twice* the usual values in the silica literature where only silicon atoms are reported [9]. For example, a 12-ring in this article contains 6 silicon and 6 oxygen atoms.

#### D. The $H_1$ barcode

The  $H_1$  barcode is proposed here as a set of intervals corresponding to algebraically independent rings in a local atomic environment and is defined in terms of the first homology group. An interval of the form  $(j, k)$  with  $0 \leq j < k$  corresponds to a ring of unknown length whose atoms are all between distance  $j$  and  $k$  of the root. An interval  $(0, k)$  more specifically corresponds to a ring containing the root atom and either  $2k - 1$  or  $2k$  total atoms. If there are fewer intervals of the form  $(0, k)$  than primitive rings passing through the root, then there is an interval of the form  $(j, k)$  which encodes information about the relationships between the primitive rings; Fig. 3 gives an example.

##### 1. Definition of homology

The definition of homology as it relates to the  $H_1$  barcode is more involved than the definitions of the other descriptors considered here; this arises from the requirement that the rings be *algebraically independent*. As motivation for the discussion below, consider that while there is no unique basis for the Euclidean plane, every basis contains precisely two vectors. That is, the space is clearly two dimensional despite the ambiguity of this choice. The situation is (perhaps surprisingly) similar for rings in a bond network. Consider the three rings in Fig. 4(b); in a sense that is made precise below, any two of these rings can be composed to generate the third in the same way that any two linearly independent vectors in the plane can be regarded as a basis. While the identity of the rings is not well defined, the number of such algebraically independent rings certainly is well defined, and the  $H_1$  barcode indicates the changes to this number as a function of  $r$ .

The first homology group is a vector space whose dimension equals the number of algebraically independent rings in a bond network (refer to Ref. [22] for an introduction to general homology theory). If  $G$  is a bond network, then the chain group  $C_0(G)$  is defined as the vector space of all formal sums  $\sum_{i=1}^k a_i v_i$ , where  $a_i \in \mathbb{R}$  and  $v_i$  is an atom of  $G$ ; this effectively attaches a real number to each atom. Similarly,  $C_1(G)$  is the vector space of all formal sums  $\sum_{i=1}^k a_i (v_i, w_i)$ , where  $a_i \in \mathbb{R}$  and  $(v_i, w_i)$  is a pair of atoms of  $G$  connected by a bond, with the relation  $(v_i, w_i) = -(w_i, v_i)$ ; this effectively attaches a real number to each *oriented* bond. If an element of  $C_1(G)$  is regarded as defining the rate of fluid flow along each bond of the network, then it is natural to ask the corresponding

rate of fluid accumulation around each atom. The linear function  $\partial : C_1(G) \rightarrow C_0(G)$  is defined by

$$\partial \left[ \sum_{i=1}^k a_i (v_i, w_i) \right] = \sum_{i=1}^k a_i v_i - a_i w_i$$

and effectively calculates this quantity for every atom of the network simultaneously. The first dimensional homology of  $G$  is then the kernel of  $\partial$  or the set of all balanced fluid flows:

$$H_1(G) = \{ \sigma \in C_1(G) : \partial(\sigma) = 0 \}.$$

$H_1(G)$  is generated by oriented rings of  $G$  or, equivalently, every balanced fluid flow can be constructed as a superposition of linearly independent flows around a well-defined number of closed circuits (though the set of closed circuits is not uniquely defined). For example, in Fig. 4(b),  $(a, b) + (b, c) + (c, d) + (d, a) \in H_1(G)$  but  $(a, b) + (b, c) + (c, d) + (a, d) \notin H_1(G)$ . Also, note that if  $\sigma_1$ ,  $\sigma_2$ , and  $\sigma_3$  are the three primitive rings

$$\sigma_1 = (a, b) + (b, c) + (c, d) + (d, a)$$

$$\sigma_2 = (a, b) + (b, e) + (e, d) + (d, a)$$

$$\sigma_3 = (b, c) + (c, d) + (d, e) + (e, b),$$

then there exists the relation

$$\sigma_1 = \sigma_2 + \sigma_3.$$

That is, as claimed in the opening of this section, only two of the three rings are algebraically independent.

The rank of the vector space  $H_1(G)$  gives the number of algebraically independent rings of  $G$ . For example, the ranks of the first homology groups of the configurations in Figs. 4(a) and 4(b) are 3 and 2, respectively. Practically speaking, the rank of  $H_1(G)$  can be efficiently computed using only information about the Euler characteristic and number of connected components of  $G$  (two atoms are in the same connected component if there is a path connecting them):

$$\begin{aligned} \chi(G) &= \text{no. of atoms} - \text{no. of bonds} \\ &= \text{no. of components} - \text{rank}[H_1(G)]. \end{aligned}$$

Solving for the rank of  $H_1(G)$  gives the concise equation

$$\begin{aligned} \text{rank}[H_1(G)] &= \text{no. of components} \\ &\quad - \text{no. of atoms} + \text{no. of bonds}. \end{aligned} \tag{1}$$

Finally, we emphasize that the vector space  $H_1(G)$  does not have a distinguished basis of “shortest” rings; each pair of primitive rings in Fig. 4(b) is a viable basis.

##### 2. Definition of the $H_1$ barcode

If  $v$  is a root atom in a bond network  $G$  and  $i \leq j$ , then the  $i, j$ -shell annulus  $S(i, j)$  at  $v$  is the subgraph composed of all atoms between distances  $i$  and  $j$  from  $v$  and the included bonds.  $F(i, j)$  is defined as the number of algebraically independent rings of  $S(i, j)$  or

$$F(i, j) = \text{rank}\{H_1[S(i, j)]\}. \tag{2}$$

For example, the local atomic environment in Fig. 5 has  $F(0, 5) = 3$ ,  $F(0, 4) = 2$ , and  $F(2, 5) = 1$ . If  $(i, j) \subseteq (k, l)$ , then rings that are algebraically independent in  $S(i, j)$  remain

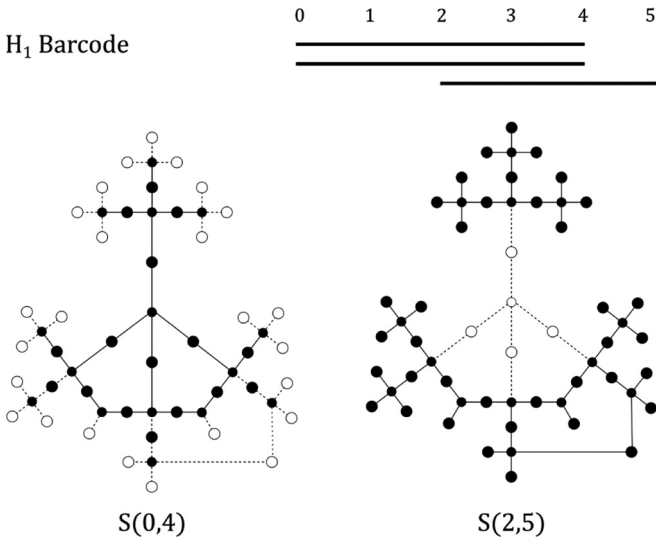


FIG. 5. Two shell annuli of a local atomic configuration and the corresponding  $H_1$  barcode.

algebraically independent in  $S(k, l)$ , implying that  $F(i, j) \leq F(k, l)$ . It follows that there is a unique set of intervals  $\mathcal{B}$  (the  $H_1$  barcode) such that

$$F(i, j) = |\{I \in \mathcal{B} : I \subseteq (i, j)\}|.$$

The information in  $\mathcal{B}$  for a root  $v$  is displayed in a barcode, as in Fig. 5 for which  $\mathcal{B} = \{(0, 4), (0, 4), (2, 5)\}$ .

The  $H_1$  barcode could alternatively be defined in terms of the zigzag persistent homology [23] or extended persistent homology [24] of the distance function to the root vertex. Those concepts could also be used to define other descriptors that contain different information about the local atomic environment. Here the  $H_1$  barcode was computed using Möbius inversion as described in Appendix A, and (perhaps surprisingly) was faster than an algorithm based on extended persistent homology.

### E. Speed of computation

The local atomic environments of a bond network are classified by iterating through a list of all possible root atoms and computing the selected descriptor for the local atomic environment around each one. The numbers of occurrences of every observed equivalence class are stored in a hash table, with an overall runtime that is linear in the number of root atoms. While the majority of the runtime is spent on computing the descriptors, the asymptotic runtime of these algorithms is not particularly relevant as the local atomic environments contain a small number of atoms. Computations were performed on a 2.3-GHz AMD Opteron processor with 64 GB of RAM, and the code was compiled with GCC version 7.3.1. The source code is open source and is available online.<sup>1</sup>

Table I shows the time required to classify  $10^5$  atomic environments at radius  $r = 6$ . The coordination profile was the fastest, requiring only 1.8 s to compute as this descriptor

TABLE I. Seconds required to classify  $10^5$  atomic environments at radius 6. The data are from a molecular dynamics simulation of glasses cooled at  $5 \times 10^{11}$  K/s, which is described in the next section.

Method	Time (s)
$H_1$ barcode	12.6
Primitive ring	23.6
Coordination	1.8

involved no additional computations once the local atomic environment was known. The  $H_1$  barcode required 12.6 s by comparison. The primitive ring profile was the slowest of the three, requiring 23.5 s in the best case. Our implementation used the Yuan and Cormack [21] primitive ring algorithm where the global structure of the bond network is used to accelerate the computation of individual profiles. Without this optimization the runtime for the primitive ring profile was 127.1 s, still substantially faster than the algorithm of Hobbs *et al.* [8] which required 476.7 s.

## III. APPLICATIONS TO SILICA

Silicon dioxide, otherwise known as silica, exists in a variety of crystalline and glassy forms at ambient conditions. Silicon atoms are generally bonded to four oxygen atoms, and oxygen atoms are generally bonded to two silicon atoms, though this is not universally true in poorly relaxed systems, and the coordination of Si atoms can be as high as 6 when under compression [25]. The four structural descriptors of local atomic environments described in Sec. II were applied to three crystalline forms of silica and to silica glasses generated by molecular dynamics simulations, quenched from the liquid at  $5 \times 10^{11}$  K/s,  $5 \times 10^{12}$  K/s, and  $5 \times 10^{13}$  K/s. An informative descriptor of local structure would ideally distinguish all of these cases. All local atomic environments are rooted at silicon atoms in the following.

The details of the molecular dynamics simulations are described in Appendix C. For each quench rate, 100 configurations of 3000 atoms (1000 Si atoms and 2000 O atoms) were generated, resulting in data sets of  $10^5$  local atomic environments rooted at silicon atoms. The Si-O bonds are defined as pairs of Si and O atoms closer than 2.2 Å, the first minimum of the radial distribution function of the silica glasses. This produces environments that are relatively stable to perturbations of the cutoff; if the cutoff is changed by  $\pm 0.2$  Å, then 98.3% of the radius 6 environments in glass quenched at a rate of  $5 \times 10^{11}$  K/s are unchanged. The corresponding percentages for glasses quenched at rates of  $5 \times 10^{12}$  K/s and  $5 \times 10^{13}$  K/s are 97.0% and 95.0%, respectively. As such, a different choice of cutoff within that range would not significantly affect the analysis below.

Coordination defects are present at low rates in all the quenched silica glasses. As expected [14], they occur most often at the highest cooling rate for which 1.47% of silicons are 5-valent, 0.96% of oxygens are 3-valent, and 0.22% of oxygens are 1-valent. Although rare from the standpoint of individual atoms, the percentage of radius 6 environments that contain at least one coordination defect is 24.2%, 41.5%, and

<sup>1</sup><https://github.com/bschweinhart/Swatches>.

TABLE II. The  $H_1$  barcode, primitive ring profile, and shell count for three crystalline forms of silica:  $\alpha$ -quartz (Q), cristobalite (C), and tridymite (T).

	$H_1$ barcode	Primitive rings	Shell count
Q	$3 \times (0, 6), 3 \times (2, 6)$	Six 12-rings	(1, 4, 4, 12, 12, 36, 30)
C	$4 \times (0, 6), 5 \times (2, 6), 4 \times (4, 6)$	Twelve 12-rings	(1, 4, 4, 12, 12, 36, 24)
T	$3 \times (0, 6), 7 \times (2, 6), 1 \times (4, 6)$	Twelve 12-rings	(1, 4, 4, 12, 12, 36, 25)

64.1% for glasses cooled at  $5 \times 10^{11}$  K/s,  $5 \times 10^{12}$  K/s, and  $5 \times 10^{13}$  K/s, respectively. Some of the descriptors considered here are more sensitive to these coordination defects than others. To illustrate this and to give a different perspective on the relative merits of the descriptors, our methodology is applied both to the full samples of  $10^5$  atomic environments at each quench rate and to the subsamples of perfectly coordinated environments.

Crystalline forms of silica are considered in Sec. III A before proceeding to the glassy structures. The Shannon entropies of the empirical probability distributions are computed in Sec. III B, indicating the extent to which the various descriptors provide informative descriptions of the local structure as a function of radius. Based on this preliminary analysis and on that of the crystalline structures, further analysis is restricted to the  $H_1$  barcode, primitive ring profile, and coordination profile at radius 6 and graph isomorphism at radius 5. Section III C uses the mutual information to compare the information provided by the different descriptors. Finally, we compare the local structure of silica glasses produced by different cooling rates in Sec. III D.

### A. Crystal structures

Table II shows the  $H_1$  barcode, primitive ring profile, and coordination profiles of local atomic environments rooted at the silicon atoms of three different crystalline forms of silica, namely  $\alpha$ -quartz, cristobalite, and tridymite (the absence of coordination defects makes the shell count equivalent to the coordination profile here). None of the descriptors is able to differentiate between the crystal structures for radii  $r \leq 5$ . The  $H_1$  barcodes and coordination profiles of cristobalite and tridymite begin to differ at radius  $r = 6$ , but the primitive ring profile is unable to distinguish these structures at any radius. This indicates that any descriptor of local atomic environments in silica should be computed at radius 6 or above to be substantially informative.

### B. Shannon entropy

An ideally informative descriptor of local atomic environments would retain enough information to differentiate environments but not so much as to regard each one as unique. Whether a descriptor is too informative can be evaluated by computing the Shannon entropy [26] of the corresponding empirical probability distribution. Recall from Sec. II that, given a bond network with atoms  $V$ , a descriptor  $X$  allows the definition of an empirical probability distribution  $P_X$  on

TABLE III. Scaled Shannon entropies of the empirical probability distributions for various descriptors for  $4 \leq r \leq 9$ . Data are computed for a sample of  $10^5$  local atomic environments in silica glasses cooled at  $5 \times 10^{11}$  K/s.

$r$	Graph iso.	P. cluster	$H_1$	P. rings	Coordination
4	0.136	0.101	0.099	0.094	0.142
5	0.445	0.358	0.270	0.226	0.299
6	0.983	0.861	0.471	0.381	0.473
7	1.000	0.999	0.704	0.550	0.678
8	1.000	—	0.902	0.706	0.846
9	1.000	—	0.986	0.821	0.965

equivalence classes  $x$ . The Shannon entropy associated with the descriptor  $X$  is defined as

$$H(X) = - \sum_x P_X(x) \log[P_X(x)].$$

$H(X)$  is minimized by a descriptor that places all atoms in a single equivalence class, resulting in an entropy of 0. On the other hand,  $H(X)$  is maximized if each atom is in an equivalence class unto itself, resulting in an entropy of

$$H(X) = - \sum_{v \in V} \frac{1}{|V|} \log(|V|^{-1}) = \log(|V|).$$

Normalizing  $H(X)$  by  $1/\log(|V|)$  gives scaled entropies between 0 and 1, with the property that  $X$  is an uninformative descriptor of local structure if the scaled entropy is close to either bound (though for different reasons).

Table III gives the scaled Shannon entropies for the sample of  $10^5$  local atomic environments in silica glasses cooled at a rate of  $5 \times 10^{11}$  K/s. In addition to graph isomorphism, the  $H_1$  barcode, the primitive ring profile, and the coordination profile, the primitive cluster of Hobbs *et al.* [8], is also calculated (truncated by the radius). Recall from the Introduction that the primitive cluster is the union of the primitive rings containing the root vertex, and that equivalence classes of primitive clusters are defined using graph isomorphism.

The scaled entropies for graph isomorphism and primitive clusters are already close to 1.0 for  $r = 6$ , indicating that they do not provide an informative classification at this radius because the empirical probability distribution cannot be constructed. This is a consequence of the fact that the number of distinct equivalence classes observed approaches the total number of environments in the sample ( $9.8 \times 10^4$  different graph isomorphism classes and  $8.6 \times 10^4$  different primitive cluster classes). On the other hand, the primitive ring profile,  $H_1$  barcode, and coordination profile have scaled entropies ranging from 0.381 to 0.473 at  $r = 6$  and provide informative classifications at this radius. The primitive ring profile can even be used for larger radii since the number of possible equivalence classes grows more slowly as a function of radius for this descriptor. Since the Shannon entropy is relatively insensitive to the tail of the probability distribution (by design), analogous data for the subsample of perfectly coordinated environments are nearly identical.

TABLE IV. The uncertainty coefficients of three classifiers at  $r = 6$  and graph isomorphism at  $r = 5$  applied to a data set of  $10^5$  local atomic environments in silica glasses cooled at  $5 \times 10^{11}$  K/s.

	$H_1$ (6)	P. rings (6)	Coord. (6)	Graph iso. (5)
$H_1$ (6)	1.00	0.52	0.83	0.62
P. rings (6)	0.65	1.00	0.62	0.63
Coordination (6)	0.83	0.50	1.00	0.61
Graph iso. (5)	0.66	0.54	0.65	1.00

### C. Mutual information

A natural question to ask is whether different descriptors encode similar information about local atomic environments. This is measured by means of the uncertainty coefficient  $U(X|Y)$  between the empirical probability distributions induced by descriptors  $X$  and  $Y$ ; the uncertainty coefficient effectively indicates the amount of information about  $P_X$  that can be deduced given prior knowledge of  $P_Y$  [27]. The precise definition involves the empirical joint distribution

$$P_{X,Y}(x, y) = |\{v \in V : X(v) = x, Y(v) = y\}|/|V|.$$

The mutual information [26] of  $P_X$  and  $P_Y$  can then be defined as

$$I(X; Y) = \sum_{x,y} P_{X,Y}(x, y) \log \left[ \frac{P_{X,Y}(x, y)}{P_X(x)P_Y(y)} \right].$$

Finally, the uncertainty coefficient between  $P_X$  and  $P_Y$  is the ratio of the mutual information to the Shannon entropy of  $P_X$ , or

$$U(X|Y) = I(X; Y)/H(X).$$

$U(X|Y) = 0$  if and only if  $P_X$  and  $P_Y$  are independent (i.e., none of the information encoded by  $X$  is given by  $Y$ ). Conversely,  $U(X|Y) = 1$  if and only if  $Y(v_1) = Y(v_2)$  implies that  $X(v_1) = X(v_2)$  for all  $v_1, v_2 \in V$  (i.e., all of the information encoded by  $X$  is given by  $Y$ ). Observe that this definition is inherently asymmetric. Table IV gives the uncertainty coefficients between the  $H_1$  barcode ( $B_6$ ), primitive ring profile ( $P_6$ ), and coordination profile ( $V_6$ ) at radius  $r = 6$ , and graph isomorphism at radius  $r = 5$ , for the data set of  $10^5$  local atomic environments in silica glasses cooled at  $5 \times 10^{11}$  K/s. The uncertainty coefficients  $U(B_6|V_6) = 0.83$  and  $U(V_6|B_6) = 0.83$  are quite high and substantially higher than  $U(B_6|P_6) = 0.52$  and  $U(V_6|P_6) = 0.50$ . This can be explained by observing that the  $H_1$  barcode and coordination profile consider all atoms in the local atomic environment of radius 6 rather than just those in the local cluster. What is perhaps surprising is that  $U(P_6|B_6) = 0.65$  is not higher. This indicates that the number and length of primitive rings at the root encodes different information than the number and length of algebraically independent rings in the local atomic environment. That the values in the fourth column are all around 0.6 indicates that a large amount of information is gained by considering atoms in the 6th neighbor shell. If the  $H_1$  barcode and primitive ring profile were instead constructed for  $r = 5$ , then they would be fully explained by graph isomorphism at that radius, resulting in uncertainty coefficients of 1.00.

TABLE V. The same data as in Table IV but for the subsample of perfectly coordinated environments.

	$H_1$ (6)	P. rings (6)	Coord. (6)	Graph iso. (5)
$H_1$ (6)	1.00	0.53	0.82	0.58
P. rings (6)	0.63	1.00	0.59	0.60
Coordination (6)	1.00	0.60	1.00	0.60
Graph iso. (5)	0.68	0.59	0.57	1.00

Table V shows the uncertainty coefficient computations for the subsample of 75 849 perfectly coordinated local atomic environments in glasses cooled at  $5 \times 10^{11}$  K/s. The most dramatic change is that  $U(V_6|B_6) = 1.00$ , indicating that the shell count provides a strictly coarser classification (or strictly less information) than the  $H_1$  barcode for perfectly coordinated environments. In fact, the shell count provides the same information as the endpoints of the  $H_1$  barcode intervals in this case, as shown in Appendix B.

### D. Different cooling rates

The initial analysis of silica glasses prepared at different cooling rates involves comparing the shapes of the empirical probability distributions for three different descriptors in Fig. 6. The rank of an equivalence class in the glass produced with the slowest quench rate is given on the  $x$  axis and the frequency for each of the three cooling rates on the  $y$  axis. The blue curve corresponding to the slowest quench rate  $5 \times 10^{11}$  K/s is monotonically decreasing, but the green and red curves corresponding to quench rates of  $5 \times 10^{12}$  K/s and  $5 \times 10^{13}$  K/s are jagged; this is because the ranking of the most common equivalence classes is different in the three cases. While the rank may change, common equivalence classes for one cooling rate tend to be common in the others as well (in the print version, the blue curve is shown in dark gray, the red curve in medium gray, and the green curve in light gray).

In all three figures, the blue curve is highest followed by the green and then the red. That is, the empirical probability distributions for glasses cooled at faster rates are broader, the most common classes are relatively less common, and more probability mass is in the tail of the distribution. In other words, glasses cooled at faster rates exhibit more disorder in their local structure. The corresponding plots for the perfectly coordinated subsamples (not shown) are similar for the  $H_1$  barcode and primitive ring profile, with slightly higher frequencies and slightly less separation between the cooling rates. The coordination profile provides substantially less separation for the subsample; this change is quantified below.

#### 1. KL divergence

The ability of the descriptors to differentiate glasses produced at different quench rates is quantified by the symmetrized Kullback-Leibler (KL) divergence [17] between the corresponding empirical probability distributions. The KL divergence  $D_{\text{KL}}(P_X \| Q_X)$  effectively measures the relative entropy of the probability distribution  $P_X$  with respect to  $Q_X$  and is perhaps the most natural extension of the Shannon entropy to this context. While the KL divergence is asymmetric, the symmetrized version conforms more to our intuitive notions

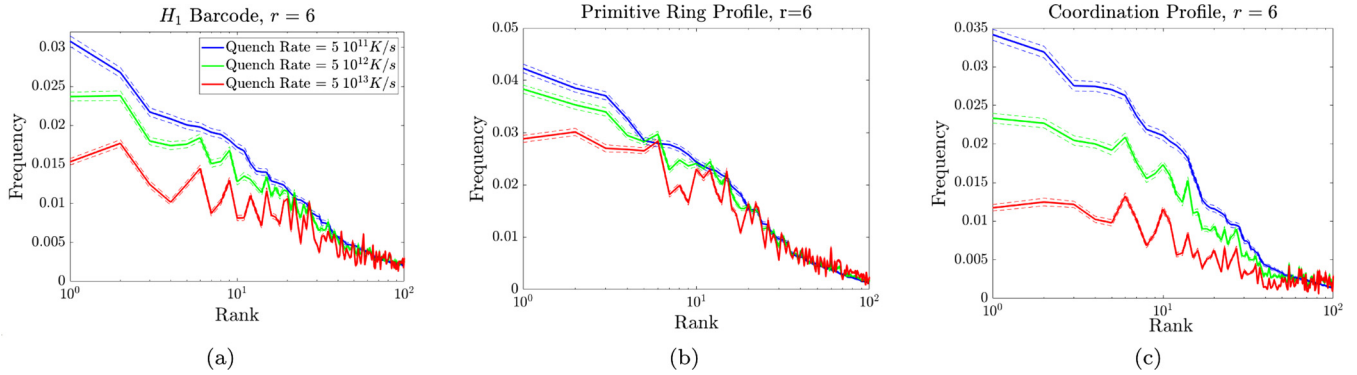


FIG. 6. Empirical probability distributions of equivalence classes in silica glasses produced at different cooling rates. The equivalence classes are ranked by frequency for the glass with the slowest cooling rate, with the frequency of the class on the y axis. The descriptors corresponding to the first 10 equivalence classes in each figure are given in Appendix D (the first column of each of the three tables). The dashed lines indicate the standard error of the frequency estimates.

of a distance (it is a semimetric on probability densities but a pseudosemimetric on bond networks).

Given a descriptor  $X$ , two bond networks  $V_1$  and  $V_2$  give the two empirical probability distributions  $P_X$  and  $Q_X$  over a single set of equivalence classes  $x$ . The KL divergence between  $P_X$  and  $Q_X$  is defined as

$$D_{\text{KL}}(P_X \| Q_X) = \sum_x P_X(x) \log \left[ \frac{P_X(x)}{Q_X(x)} \right].$$

This is symmetrized by adding the relative entropy of  $P_X$  with respect to  $Q_X$  to that of  $Q_X$  with respect to  $P_X$ , or

$$D_{\text{KL}}(P_X, Q_X) = D_{\text{KL}}(P_X \| Q_X) + D_{\text{KL}}(Q_X \| P_X).$$

Practical calculation of  $D_{\text{KL}}(P_X, Q_X)$  requires the substitution  $q \log(q/p) = 0$  if either  $q$  or  $p$  is zero and is justified under two reasonable assumptions: (1) the frequency of an equivalence class in the underlying probability distributions is nonzero in one preparation if and only if it is nonzero in the other, and (2) if an equivalence class is so rare in one preparation so as to not appear in the sample, it is also rare in the other.

Table VI shows the symmetrized KL divergences between the empirical probability densities, effectively measuring the separation between the curves in Fig. 6. Lower values indicate that larger sample sizes are required to reliably differentiate glasses produced at different cooling rates. The coordination profile provides the best differentiation between different cooling rates, followed by the  $H_1$  barcode and finally the primitive ring profile. The coordination profile’s advantage is, however, attributed to its sensitivity to coordination defects; consider Table VII, which shows the corresponding data for

TABLE VI. The symmetrized Kullback-Leibler divergence for three different descriptors with  $r = 6$  between empirical probability distributions at different cooling rates.

Comparison	$H_1$ barcode	P. rings	Coord.
$5 \times 10^{11}, 5 \times 10^{12}$	0.065	0.026	0.162
$5 \times 10^{11}, 5 \times 10^{13}$	0.288	0.146	0.655
$5 \times 10^{12}, 5 \times 10^{13}$	0.149	0.066	0.279

the subsamples of perfectly coordinated environments. In this case, the  $H_1$  barcode provides the best discrimination between quench rates, followed by the coordination profile, and finally the primitive ring profile.

Symmetrized KL divergences were also computed between the empirical probability densities for the full sample of  $10^5$  local atomic environments in glasses cooled at  $5 \times 10^{13}$  K/s and the subsample of  $3.6 \times 10^4$  perfectly coordinated environments. The divergence for the coordination profile is very large (0.657), and much larger than those given by the  $H_1$  barcode (0.086) or the primitive ring profile (0.050). That is, the coordination profile is quite sensitive to coordination defects, but the other two descriptors are not. Divergences for glasses cooled at slower rates were smaller since perfectly coordinated environments made up a larger proportion of the total sample.

### 2. Ranked examples

We consider examples of the most common equivalence classes and the ones that are relatively over-represented in glasses produced with a faster cooling rate. At radius 5, Figure 7 shows the six most common graph isomorphism classes in glasses produced at a cooling rate of  $5 \times 10^{11}$  K/s, with the corresponding barcode, primitive ring profile, and shell count (there were no coordination errors in these examples).  $f_1$  and  $r_1$  are the frequency and rank of the isomorphism class at a cooling rate of  $5 \times 10^{11}$  K/s, and  $f_2$  and  $r_2$  are the corresponding quantities for a cooling rate of  $5 \times 10^{13}$  K/s. The ranking is very similar for both conditions, reinforcing that the set of most frequent classes is relatively insensitive to cooling rate. As expected, environments with 10-rings

TABLE VII. The same data as in Table VII but for the subsamples of perfectly coordinated environments (for which the coordination profile gives equivalent information to the shell count).

Comparison	$H_1$ barcode	P. rings	Coord.
$5 \times 10^{11}, 5 \times 10^{12}$	0.044	0.019	0.021
$5 \times 10^{11}, 5 \times 10^{13}$	0.198	0.121	0.140
$5 \times 10^{12}, 5 \times 10^{13}$	0.112	0.062	0.069



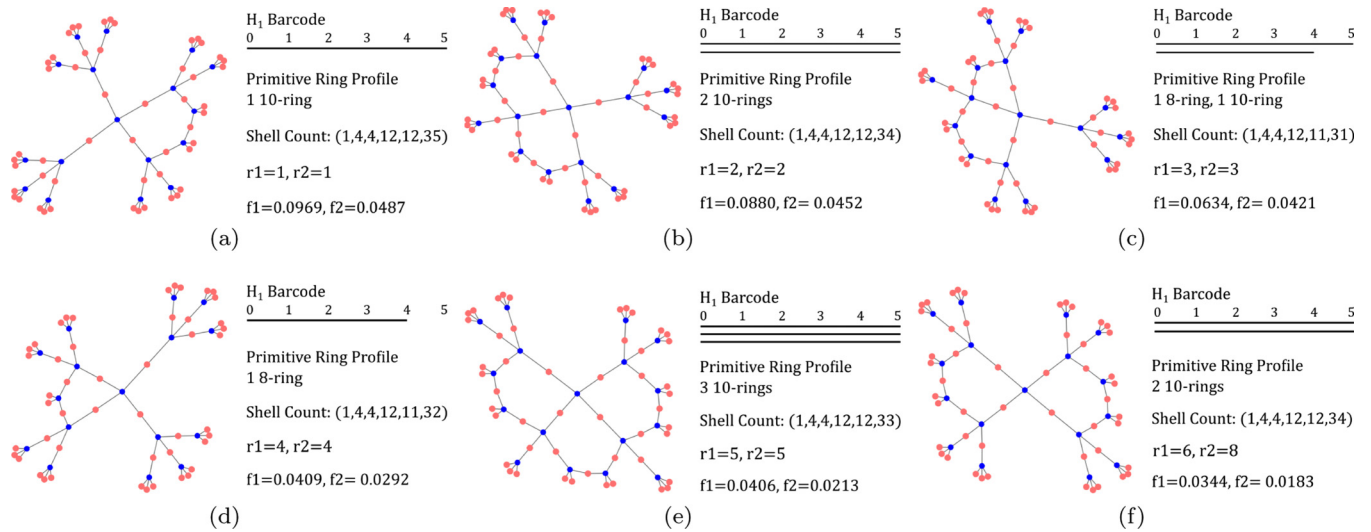


FIG. 7. The six most frequent graph isomorphism classes at radius 5 in glasses cooled at a rate of  $5 \times 10^{11}$  K/s. Graph embeddings were generated using a force-directed layout [28]. Silicon atoms are shown in blue (dark gray), and oxygen atoms are shown in red (light gray).

are well represented, but Figs. 7(c) and 7(d) have 8-rings. More interestingly, several of the most common environments include adjacent rings, i.e., an Si-O-Si bond which is shared by two or three 10-rings as in Figs. 7(b) and 7(d) or by a 10- and an 8-ring as in Fig. 7(c). The adjacency of these rings cannot be detected by ring statistics alone.

Figure 8 shows the six graph isomorphism types that maximize  $f_2 - f_1$ , i.e., the types most over-represented in glasses produced with the fastest cooling rate. The data reveal that glasses produced with a faster cooling rate have relatively more short rings with six or eight atoms, and configurations with adjacent short rings are especially favored. This is consistent with the fact that, at lower quench rates, the number of 12 rings (with 6 Si atoms) increases while that of other rings decreases [14].

Tables VIII–X in Appendix D show similar data for the  $H_1$  barcode, primitive ring profile, and coordination profile at

radius  $r = 6$ . For each descriptor, the top 10 classes ranked by  $f_1$ ,  $f_1 - f_2$ , and  $f_2 - f_1$  are listed, using the same notation as in the previous paragraph. The descriptors that maximize  $f_2 - f_1$  correspond to the peaks in Fig. 6 where the red (medium gray) curve is above the blue (dark gray) one. Examining Table VIII or IX reveals that local environments in glasses produced at  $5 \times 10^{11}$  K/s have more rings with 10 or 12 atoms (five or six Si atoms), while a cooling rate of  $5 \times 10^{13}$  K/s yields more 6- and 8-rings (three or four Si atoms). For example, the early peak in Fig. 6(b) corresponds to a primitive ring profile with one primitive 8-ring, two primitive 10-rings, and two primitive 12-rings. Moreover, glasses produced by faster cooling rates exhibit more coordination errors. In Table X, all of the top 10 coordination profiles ranked by  $f_2 - f_1$  include a silicon in neighbor shell 6 that is adjacent to five oxygen atoms, and one contains an oxygen adjacent to three silicon atoms.

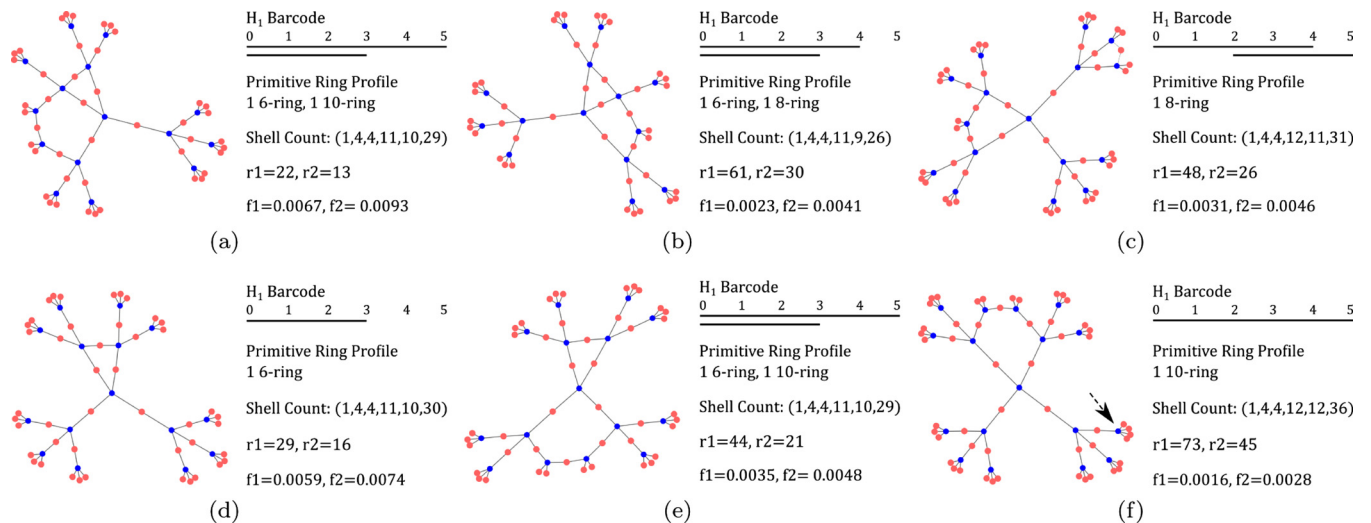


FIG. 8. The six graph isomorphism classes at radius 5 that maximize  $f_2 - f_1$ . Silicon atoms are shown in blue (dark gray), and oxygen atoms are shown in red (light gray).

#### IV. CONCLUSION

The connectivity of the bonds in a network solid is believed to be involved in a variety of physical quantities, e.g., the viscosity and configurational entropy of covalent glass forming materials. The authors suggest that part of the reason that the dependence of the configurational entropy on composition cannot yet be reliably predicted is the lack of a standard approach to quantify the change to the network connectivity with a change in atomic valence. Defining a probability distribution of local atomic environments that is (by hypothesis) characteristic of identically prepared bond networks establishes a base that can be used to address this issue and has the further advantage of unifying the treatment of crystalline and amorphous materials.

A closely related question, and one that is the main subject of this article, is precisely which type of information should be used to classify local atomic environments. Strictly ascribing to the principle that more information is always preferable leads to the situation that all geometric and topological information about the local atomic environments is retained, rendering the approximation of the underlying probability distribution by any finite sampling untenable. Retaining only a subset of the information is therefore a practical necessity, and depending on precisely what information is retained could actually serve to clarify the relationship of the material's structure and properties. For example, the vibrational entropy of a network solid is certainly more closely related to the number of covalent bonds and the constraints they impose than to the precise geometric arrangement of the atoms.

Given this background, four descriptors of local atomic environments were described and applied to silica networks. An appropriately informative descriptor should not only be able to distinguish the different crystalline forms of silica but should also be able to differentiate silica glasses produced at different cooling rates on the basis of variations in the bond network connectivity. Of the descriptors considered here, the coordination profile and  $H_1$  barcode at radius 6 performed best at these tasks and were also faster to compute than the primitive ring profile at the same radius. The efficacy of the coordination profile could be attributed to its sensitivity to coordination defects though; when the subset of local atomic environments without these defects was considered, the  $H_1$  barcode outperformed all the other descriptors. Moreover, the  $H_1$  barcode and the primitive ring profile could be more readily interpreted than the coordination profile in terms of ring statistics. Overall, the  $H_1$  barcode appears to be an efficient, informative, and interpretable descriptor for local atomic environments in silica and could be invaluable to further advances in our understanding of the structure of amorphous materials.

#### ACKNOWLEDGMENTS

We thank D. Morozov for pointing out the connection between the  $H_1$  barcode and extended or zigzag persistent homology. B. Schweinhart was partially supported by NSF Mathematical Sciences Postdoctoral Research Fellowship under Award No. DMS-1606259, J.K.M. was supported by the National Science Foundation under Grant No. DMR 1839370,

and D.R. acknowledges support from LABEX iMUST (ANR-10-LABX-0064) of Université de Lyon (programme Investissements d'Avenir, ANR-11-IDEX-0007).

#### APPENDIX A: COMPUTATION OF THE $H_1$ BARCODE

The  $H_1$  barcode is computed using Möbius inversion [29], which is a generalized version of the inclusion-exclusion principle for partially ordered sets (Ref. [30] gives an introduction).

Let  $S$  be the set of all subintervals  $(a, b)$  of  $(0, r)$ , and let  $(a, b) \leq (c, d)$  if  $(a, b) \subseteq (c, d)$ . The Möbius function of  $S$  is defined recursively for all  $(a, b) \leq (c, d)$  by

$$\begin{aligned}\mu[(a, b), (a, b)] &= 1 \\ \mu[(a, b), (c, d)] &= - \sum_{(a,b) \leq (e,f) \leq (c,d)} \mu[(a, b), (e, f)].\end{aligned}$$

It is a mathematical theorem [29] that if  $F$  is a function of the form

$$F[(c, d)] = \sum_{(a,b) \leq (c,d)} G[(a, b)],$$

then  $G$  can be recovered by the formula

$$G[(c, d)] = \sum_{(a,b) \leq (c,d)} F[(a, b)] \mu[(a, b), (c, d)].$$

In particular, if  $F[(a, b)]$  is given as in Eq. (2), then  $G[(a, b)]$  is the number of intervals of the form  $(a, b)$  in the  $H_1$  barcode. Practically,  $F[(a, b)]$  is computed using the Euler characteristic as in Eq. (1) before applying the previous formula to compute the number of intervals.

#### APPENDIX B: THE $H_1$ BARCODE OF PERFECTLY COORDINATED ENVIRONMENTS

We show that the information in the shell count is equivalent to that of the endpoints of the  $H_1$  barcode intervals for perfectly coordinated environments, where a perfectly coordinated environment is a bipartite rooted graph so vertices in even shells have degree  $d_0$  and vertices in odd shells have degree  $d_1$ . For the perfectly coordinated silica environments considered in the text,  $d_0 = 4$  and  $d_1 = 2$ .

*Proposition 1.* Let  $G$  be a perfectly coordinated local atomic environment of radius  $r$ . The endpoints of the  $H_1$  barcode intervals of  $G$  can be computed in terms of the shell count and vice versa.

*Proof.* First, note that the number of  $H_1$  barcode intervals whose endpoints are  $\leq r_0$  equals (by definition)  $F(0, r_0)$ , the rank of the first homology group of the local atomic environment of radius  $r_0$ . As such, knowledge of the endpoints is equivalent to knowledge of  $F(0, r_0)$  for all  $r_0 \leq r$ . Assume that we know  $F(0, r_0)$  for all  $r_0 \leq r$  and suppose by induction that we have computed the shell count at radius  $r_0$  [ $A(r_0)$ ] as well as the number of bonds between the atoms in shells  $r_0$  and  $r_0 - 1$  [ $B(r_0, r_0 - 1)$ ] for all  $r_0 < r$ . The environment is bipartite, so atoms in shell  $r - 1$  can share bonds with atoms in shells  $r$  or  $r - 2$  but not atoms in shell  $r - 1$ . It follows that

$$B(r, r - 1) = dA(r - 1) - B(r - 1, r - 2),$$

where  $d = d_0$  if  $r - 1$  is even and  $d = d_1$  otherwise. The number of bonds of  $G$  equals

$$B(G) = \sum_{r_0 \leq r} B(r_0, r_0 - 1),$$

and we can use Eq. (1) to find the number of atoms of  $G$ :

$$A(G) = 1 + B(G) - F(0, r_0).$$

The shell count at radius  $r$  is then

$$A(r) = A(G) - \sum_{r_0 < r} A(r_0).$$

A similar argument shows that we can compute  $F(0, r_0)$  for all  $r_0 \leq r$  given knowledge of the shell counts. ■

**APPENDIX C: MOLECULAR DYNAMICS METHODOLOGY**

Molecular dynamics simulations of silica glasses were performed using the classical two-body potential developed by

van Beest, Kramer, and van Santen [31], modified by Carreé *et al.* [32] to replace the long-range Coulombic interaction by a Wolf truncation. The parameters of the potential are given in Ref. [33]. The glasses considered here contained 3000 atoms (1000 Si and 2000 O atoms) in cubic simulation cells with periodic boundary conditions and a density of 2.2 g/cm<sup>3</sup>. They were produced from melts equilibrated at 5200 K and quenched at constant volume to 10 K at rates between  $5 \times 10^{11}$  K/s and  $5 \times 10^{13}$  K/s. The simulations used a time step of 1 fs and an Andersen thermostat [34] which reassigned atomic velocities with a probability of 0.001 per time step. The quenches were followed by energy minimizations to obtain inherent structures before further analyses.

**APPENDIX D: RANKED EXAMPLES**

Tables VIII–X show the  $H_1$  barcode, primitive ring profile, and coordination profile classes that maximize  $f_1$ ,  $f_1 - f_2$ , and  $f_2 - f_1$ , where  $f_1$  is the frequency of the class in glasses cooled at a rate of  $5 \times 10^{11}$  K/s, and  $f_2$  is the corresponding quantity for glasses cooled at a rate of  $5 \times 10^{13}$  K/s. See Sec. III D 2 for details and analysis.

TABLE VIII. The  $H_1$  barcodes at radius 6 that maximize  $f_1$ ,  $f_1 - f_2$ , and  $f_2 - f_1$ .  $f_1$  and  $r_1$  are the frequency and rank of the isomorphism class at a cooling rate of  $5 \times 10^{11}$  K/s, and  $f_2$  and  $r_2$  are the corresponding quantities for a cooling rate of  $5 \times 10^{13}$  K/s. See Sec. III D 2 for details and analysis.

Sorted by	$f_1$	$f_1 - f_2$	$f_2 - f_1$
1	$2 \times (0, 5), (0, 6), 4 \times (2, 6)$ $f_1 = 0.033, f_2 = 0.017$ $r_1 = 1, r_2 = 2$	$2 \times (0, 5), (0, 6), 4 \times (2, 6)$ $f_1 = 0.033, f_2 = 0.017$ $r_1 = 1, r_2 = 2$	$(0, 4), (0, 5), (0, 6), (2, 5), 4 \times (2, 6)$ $f_1 = 0.004, f_2 = 0.006$ $r_1 = 58, r_2 = 31$
2	$2 \times (0, 5), (0, 6), 3 \times (2, 6)$ $f_1 = 0.031, f_2 = 0.015$ $r_1 = 2, r_2 = 3$	$2 \times (0, 5), (0, 6), 3 \times (2, 6)$ $f_1 = 0.031, f_2 = 0.015$ $r_1 = 2, r_2 = 3$	$2 \times (0, 4), (0, 5), (2, 5), 2 \times (2, 6)$ $f_1 = 0.003, f_2 = 0.005$ $r_1 = 66, r_2 = 43$
3	$(0, 4), (0, 5), (0, 6), 3 \times (2, 6)$ $f_1 = 0.027, f_2 = 0.018$ $r_1 = 3, r_2 = 1$	$(0, 5), 2 \times (0, 6), 4 \times (2, 6)$ $f_1 = 0.021, f_2 = 0.010$ $r_1 = 5, r_2 = 15$	$2 \times (0, 4), (0, 5), (2, 5), 3 \times (2, 6)$ $f_1 = 0.002, f_2 = 0.003$ $r_1 = 103, r_2 = 61$
4	$2 \times (0, 5), (0, 6), 5 \times (2, 6)$ $f_1 = 0.022, f_2 = 0.012$ $r_1 = 4, r_2 = 7$	$(0, 5), 2 \times (0, 6), 3 \times (2, 6)$ $f_1 = 0.019, f_2 = 0.009$ $r_1 = 8, r_2 = 17$	$(0, 4), (0, 5), (0, 6), 4 \times (2, 6), (4, 6)$ $f_1 = 0.003, f_2 = 0.004$ $r_1 = 78, r_2 = 50$
5	$(0, 5), 2 \times (0, 6), 4 \times (2, 6)$ $f_1 = 0.021, f_2 = 0.010$ $r_1 = 5, r_2 = 15$	$2 \times (0, 5), (0, 6), 5 \times (2, 6)$ $f_1 = 0.022, f_2 = 0.012$ $r_1 = 4, r_2 = 7$	$(0, 4), 2 \times (0, 5), 2 \times (2, 5), 3 \times (2, 6)$ $f_1 = 0.002, f_2 = 0.003$ $r_1 = 118, r_2 = 76$
6	$(0, 4), (0, 5), (0, 6), 2 \times (2, 6)$ $f_1 = 0.020, f_2 = 0.012$ $r_1 = 6, r_2 = 6$	$2 \times (0, 5), (0, 6), 2 \times (2, 6)$ $f_1 = 0.017, f_2 = 0.008$ $r_1 = 11, r_2 = 22$	$(0, 4), 2 \times (0, 5), (2, 5), 3 \times (2, 6), (4, 6)$ $f_1 = 0.001, f_2 = 0.002$ $r_1 = 151, r_2 = 93$
7	$(0, 4), (0, 5), (0, 6), 4 \times (2, 6)$ $f_1 = 0.020, f_2 = 0.014$ $r_1 = 7, r_2 = 4$	$(0, 4), (0, 5), (0, 6), 3 \times (2, 6)$ $f_1 = 0.027, f_2 = 0.018$ $r_1 = 3, r_2 = 1$	$(0, 4), 2 \times (0, 5), (2, 5), 4 \times (2, 6), (4, 6)$ $f_1 = 0.001, f_2 = 0.002$ $r_1 = 224, r_2 = 117$
8	$(0, 5), 2 \times (0, 6), 3 \times (2, 6)$ $f_1 = 0.019, f_2 = 0.009$ $r_1 = 8, r_2 = 17$	$3 \times (0, 5), 4 \times (2, 6)$ $f_1 = 0.019, f_2 = 0.010$ $r_1 = 9, r_2 = 12$	$(0, 4), 2 \times (0, 5), (2, 5), 2 \times (2, 6), (4, 6)$ $f_1 = 0.001, f_2 = 0.002$ $r_1 = 145, r_2 = 94$
9	$3 \times (0, 5), 4 \times (2, 6)$ $f_1 = 0.019, f_2 = 0.010$ $r_1 = 9, r_2 = 12$	$3 \times (0, 5), 3 \times (2, 6)$ $f_1 = 0.017, f_2 = 0.008$ $r_1 = 12, r_2 = 21$	$(0, 3), (0, 5), (0, 6), 3 \times (2, 6)$ $f_1 = 0.002, f_2 = 0.004$ $r_1 = 83, r_2 = 56$
10	$(0, 4), 2 \times (0, 5), 3 \times (2, 6)$ $f_1 = 0.018, f_2 = 0.013$ $r_1 = 10, r_2 = 5$	$(0, 4), (0, 5), (0, 6), 2 \times (2, 6)$ $f_1 = 0.020, f_2 = 0.012$ $r_1 = 6, r_2 = 6$	$2 \times (0, 4), (0, 5), 2 \times (2, 5), 2 \times (2, 6)$ $f_1 = 0.001, f_2 = 0.002$ $r_1 = 237, r_2 = 143$

TABLE IX. The primitive ring profiles at radius 6 that maximize  $f_1$ ,  $f_1 - f_2$ , and  $f_2 - f_1$ .

Sorted by	$f_1$	$f_1 - f_2$	$f_2 - f_1$
1	Two 10-rings, three 12-rings $f_1 = 0.049, f_2 = 0.035$ $r_1 = 1, r_2 = 1$	Two 10-rings, three 12-rings $f_1 = 0.049, f_2 = 0.035$ $r_1 = 1, r_2 = 1$	One 6-ring, one 10-ring, two 12-rings $f_1 = 0.004, f_2 = 0.008$ $r_1 = 61, r_2 = 33$
2	Two 10-rings, four 12-rings $f_1 = 0.042, f_2 = 0.029$ $r_1 = 2, r_2 = 3$	Two 10-rings, four 12-rings $f_1 = 0.042, f_2 = 0.029$ $r_1 = 2, r_2 = 3$	Two 8-rings, one 10-ring, one 12-ring $f_1 = 0.005, f_2 = 0.008$ $r_1 = 46, r_2 = 38$
3	Three 10-rings, two 12-rings $f_1 = 0.039, f_2 = 0.030$ $r_1 = 3, r_2 = 2$	Three 10-rings, three 12-rings $f_1 = 0.037, f_2 = 0.027$ $r_1 = 4, r_2 = 5$	One 6-ring, one 10-ring, one 12-ring $f_1 = 0.003, f_2 = 0.006$ $r_1 = 67, r_2 = 48$
4	Three 10-rings, three 12-rings $f_1 = 0.037, f_2 = 0.027$ $r_1 = 4, r_2 = 5$	Two 10-rings, five 12-rings $f_1 = 0.028, f_2 = 0.018$ $r_1 = 8, r_2 = 14$	One 6-ring, one 10-ring, three 12-rings $f_1 = 0.003, f_2 = 0.005$ $r_1 = 68, r_2 = 53$
5	Two 10-rings, two 12-rings $f_1 = 0.033, f_2 = 0.027$ $r_1 = 5, r_2 = 6$	One 10-ring, five 12-rings $f_1 = 0.026, f_2 = 0.017$ $r_1 = 10, r_2 = 16$	Two 8-rings, one 10-ring, two 12-rings $f_1 = 0.006, f_2 = 0.008$ $r_1 = 44, r_2 = 35$
6	One 8-ring, one 10-ring, three 12-rings $f_1 = 0.028, f_2 = 0.027$ $r_1 = 6, r_2 = 7$	Three 10-rings, two 12-rings $f_1 = 0.039, f_2 = 0.030$ $r_1 = 3, r_2 = 2$	One 6-ring, two 10-rings, two 12-rings $f_1 = 0.002, f_2 = 0.005$ $r_1 = 75, r_2 = 59$
7	One 8-ring, two 10-rings, two 12-rings $f_1 = 0.028, f_2 = 0.028$ $r_1 = 7, r_2 = 4$	One 10-ring, six 12-rings $f_1 = 0.018, f_2 = 0.011$ $r_1 = 19, r_2 = 26$	One 6-ring, two 10-rings, one 12-ring $f_1 = 0.003, f_2 = 0.005$ $r_1 = 69, r_2 = 55$
8	Two 10-rings, five 12-rings $f_1 = 0.028, f_2 = 0.018$ $r_1 = 8, r_2 = 14$	One 10-ring, four 12-rings $f_1 = 0.027, f_2 = 0.020$ $r_1 = 9, r_2 = 12$	Two 8-rings, two 10-rings, two 12-rings $f_1 = 0.003, f_2 = 0.005$ $r_1 = 66, r_2 = 54$
9	One 10-ring, four 12-rings $f_1 = 0.027, f_2 = 0.020$ $r_1 = 9, r_2 = 12$	Two 10-rings, two 12-rings $f_1 = 0.033, f_2 = 0.027$ $r_1 = 5, r_2 = 6$	One 6-ring, one 8-ring, one 10-ring, one 12-ring $f_1 = 0.001, f_2 = 0.003$ $r_1 = 97, r_2 = 73$
10	One 10-ring, five 12-rings $f_1 = 0.026, f_2 = 0.017$ $r_1 = 10, r_2 = 16$	Three 10-rings, four 12-rings $f_1 = 0.022, f_2 = 0.017$ $r_1 = 15, r_2 = 15$	One 6-ring, one 8-ring, one 10-ring, two 12-rings $f_1 = 0.001, f_2 = 0.003$ $r_1 = 117, r_2 = 93$

TABLE X. The coordination profiles at radius 6 that maximize  $f_1$ ,  $f_1 - f_2$ , and  $f_2 - f_1$ . The shell count is shown, and \* indicates the presence of one atom of valence 5.

Sorted by	$f_1$	$f_1 - f_2$	$f_2 - f_1$
1	(1, 4, 4, 12, 12, 34, 27) $f_1 = 0.038, f_2 = 0.014$ $r_1 = 1, r_2 = 1$	(1, 4, 4, 12, 12, 34, 27) $f_1 = 0.038, f_2 = 0.014$ $r_1 = 1, r_2 = 1$	(1, 4, 4, 12, 11, 30, 24*) $f_1 = 0.002, f_2 = 0.003$ $r_1 = 91, r_2 = 38$
2	(1, 4, 4, 12, 12, 34, 28) $f_1 = 0.034, f_2 = 0.012$ $r_1 = 2, r_2 = 5$	(1, 4, 4, 12, 12, 34, 28) $f_1 = 0.034, f_2 = 0.012$ $r_1 = 2, r_2 = 5$	(1, 4, 4, 12, 11, 30, 25*) $f_1 = 0.002, f_2 = 0.004$ $r_1 = 84, r_2 = 34$
3	(1, 4, 4, 12, 12, 33, 26) $f_1 = 0.032, f_2 = 0.012$ $r_1 = 3, r_2 = 3$	(1, 4, 4, 12, 12, 33, 26) $f_1 = 0.032, f_2 = 0.012$ $r_1 = 3, r_2 = 3$	(1, 4, 4, 12, 11, 31, 26*) $f_1 = 0.002, f_2 = 0.004$ $r_1 = 82, r_2 = 37$
4	(1, 4, 4, 12, 11, 31, 26) $f_1 = 0.028, f_2 = 0.012$ $r_1 = 4, r_2 = 4$	(1, 4, 4, 12, 12, 33, 27) $f_1 = 0.027, f_2 = 0.010$ $r_1 = 6, r_2 = 10$	(1, 4, 4, 12, 11, 30, 26*) $f_1 = 0.001, f_2 = 0.003$ $r_1 = 101, r_2 = 47$
5	(1, 4, 4, 12, 12, 34, 26) $f_1 = 0.027, f_2 = 0.010$ $r_1 = 5, r_2 = 7$	(1, 4, 4, 12, 12, 34, 26) $f_1 = 0.027, f_2 = 0.010$ $r_1 = 5, r_2 = 7$	(1, 4, 4, 12, 12, 33, 25*) $f_1 = 0.001, f_2 = 0.003$ $r_1 = 96, r_2 = 45$
6	(1, 4, 4, 12, 12, 33, 27) $f_1 = 0.027, f_2 = 0.010$ $r_1 = 6, r_2 = 10$	(1, 4, 4, 12, 11, 31, 26) $f_1 = 0.028, f_2 = 0.012$ $r_1 = 4, r_2 = 4$	(1, 4, 4, 12, 12, 34, 27*) $f_1 = 0.003, f_2 = 0.004$ $r_1 = 56, r_2 = 27$
7	(1, 4, 4, 12, 11, 30, 25) $f_1 = 0.026, f_2 = 0.013$ $r_1 = 7, r_2 = 2$	(1, 4, 4, 12, 12, 35, 28) $f_1 = 0.022, f_2 = 0.007$ $r_1 = 9, r_2 = 13$	(1, 4, 4, 12, 12, 33, 26*) $f_1 = 0.003, f_2 = 0.004$ $r_1 = 66, r_2 = 32$

TABLE X. (Continued.)

Sorted by	$f_1$	$f_1 - f_2$	$f_2 - f_1$
8	(1, 4, 4, 12, 12, 33, 25) $f_1 = 0.024, f_2 = 0.010$ $r_1 = 8, r_2 = 8$	(1, 4, 4, 12, 12, 35, 29) $f_1 = 0.020, f_2 = 0.006$ $r_1 = 13, r_2 = 19$	(1, 4, 4, 12, 11, 29, 24*) $f_1 = 0.001, f_2 = 0.002$ $r_1 = 167, r_2 = 73$
9	(1, 4, 4, 12, 12, 35, 28) $f_1 = 0.022, f_2 = 0.007$ $r_1 = 9, r_2 = 13$	(1, 4, 4, 12, 12, 33, 25) $f_1 = 0.024, f_2 = 0.010$ $r_1 = 8, r_2 = 8$	(1, 4, 4, 12, 12, 34, 28*) $f_1 = 0.003, f_2 = 0.004$ $r_1 = 64, r_2 = 31$
10	(1, 4, 4, 12, 11, 31, 27) $f_1 = 0.021, f_2 = 0.009$ $r_1 = 10, r_2 = 11$	(1, 4, 4, 12, 12, 34, 29) $f_1 = 0.019, f_2 = 0.006$ $r_1 = 14, r_2 = 18$	(1, 4, 4, 12, 11, 31, 25*) $f_1 = 0.002, f_2 = 0.003$ $r_1 = 92, r_2 = 49$

- [1] K. Y. Szeto and J. Villain, *Phys. Rev. B* **36**, 4715 (1987).
- [2] D. P. Shoemaker and C. B. Shoemaker, *Phys. Rev. B* **38**, 6319 (1988).
- [3] H. Sheng, W. Luo, F. Alamgir, J. Bai, and E. Ma, *Nature* **439**, 419 (2006).
- [4] C. L. Henley, *Phys. Rev. B* **34**, 797 (1986).
- [5] K. Y. Szeto, X. Fu, and W. Y. Tam, *Phys. Rev. Lett.* **88**, 138302 (2002).
- [6] J. K. Mason, E. A. Lazar, R. D. MacPherson, and D. J. Srolovitz, *Phys. Rev. E* **86**, 051128 (2012).
- [7] B. Schweinhart, J. K. Mason, and R. D. MacPherson, *Phys. Rev. E* **93**, 062111 (2016).
- [8] L. W. Hobbs, C. E. Jesurum, V. Pulim, and B. Berger, *Philos. Mag. A* **78**, 679 (1998).
- [9] S. V. King, *Nature* **213**, 1112 (1967).
- [10] S. Le Roux and P. Jund, *Comp. Mater. Sci.* **49**, 70 (2010).
- [11] Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escolar, K. Matsue, and Y. Nishiura, *Proc. Natl. Acad. Sci. USA* **113**, 7035 (2016).
- [12] L. Deng and J. Du, *J. Chem. Phys.* **148**, 024504 (2018).
- [13] X. Li, W. Song, K. Yang, N. A. Krishnan, B. Wang, M. M. Smedskjaer, J. C. Mauro, G. Sant, M. Balonis, and M. Bauchy, *J. Chem. Phys.* **147**, 074501 (2017).
- [14] P. Koziatek, J. Barrat, and D. Rodney, *J. Non-Cryst. Solids* **414**, 7 (2015).
- [15] A. Tilocca, *J. Chem. Phys.* **139**, 114501 (2013).
- [16] K. Vollmayr, W. Kob, and K. Binder, *Phys. Rev. B* **54**, 15808 (1996).
- [17] S. Kullback and R. A. Leibler, *Ann. Math. Stat.* **22**, 79 (1951).
- [18] B. D. McKay and A. Piperno, *J. Symb. Comput.* **60**, 94 (2014).
- [19] C. S. Marians and L. W. Hobbs, *J. Non-Cryst. Solids* **124**, 242 (1990).
- [20] L. Guttman, *J. Non-Cryst. Solids* **116**, 145 (1990).
- [21] X. Yuan and A. Cormack, *Comp. Mater. Sci.* **24**, 343 (2002).
- [22] A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, UK, 2002).
- [23] G. Carlsson, V. De Silva, and D. Morozov, in *Proceedings of the 25th Annual Symposium on Computational Geometry (ACM, New York, 2009)*, pp. 247–256.
- [24] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, *Found. Comput. Math.* **9**, 79 (2009).
- [25] C. J. Benmore, E. Soignard, S. A. Amin, M. Guthrie, S. D. Shastri, P. L. Lee, and J. L. Yarger, *Phys. Rev. B* **81**, 054105 (2010).
- [26] C. E. Shannon, *Bell Labs Tech. J.* **27**, 379 (1948).
- [27] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing* (Cambridge University Press, Cambridge, UK, 2007).
- [28] T. M. Fruchterman and E. M. Reingold, *Softw. Pract. Exp.* **21**, 1129 (1991).
- [29] G.-C. Rota, *Z. Wahrscheinlichkeitstheorie* **2**, 340 (1964).
- [30] E. A. Bender and J. R. Goldman, *Am. Math. Mon.* **82**, 789 (1975).
- [31] B. W. H. van Beest, G. J. Kramer, and R. A. van Santen, *Phys. Rev. Lett.* **64**, 1955 (1990).
- [32] A. Carré, L. Berthier, J. Horbach, S. Ispas, and W. Kob, *J. Chem. Phys.* **127**, 114512 (2007).
- [33] B. Mantsi, A. Tanguy, G. Kermouche, and E. Barthel, *Eur. Phys. J. B* **85**, 304 (2012).
- [34] H. C. Andersen, *J. Chem. Phys.* **72**, 2384 (1980).