

A Supplementary Text

A.1 Evaluating generative models for amorphous materials

Good generative models for amorphous materials should be able to reproduce quantitative distributions of structural features and properties for these systems. Traditional validation strategies in amorphous materials rely on coordination environments, order parameters, energy distributions, or similar structural features that can be easily computed from the structure. Deep learning methods often fall short of producing physically reasonable amorphous structures compared to simulation results — exhibiting unphysical coordination environments, large deviations from expected pair correlation functions, or unrealistic defects — let alone reproducing their properties or accounting for processing parameters such as cooling rates.^[18] As examples of recent work related to amorphous materials, Yang *et al.* presented a forward predictive model and an inverse generative approach for liquid electrolyte compositions,^[44] though not all tasks are performed at the atomistic scale. Recently, Chen *et al.* used a graph variational autoencoder, with energy and radial distribution function (RDF) regularizations, to generate glass structures but only validate their generations with energy distributions and RDFs.^[14] Furthermore, qualitative visualizations or distance/descriptor distributions are insufficient to prove whether generated structures are simultaneously accurate, novel, and valid. The work from Yong *et al.*^[18] thoroughly benchmarks recent generative models’ ability to produce realistic samples of amorphous and disordered systems, highlighting that most models cannot produce amorphous structures even at the qualitative level. In that work, structural characteristics and properties relevant to the analysis of amorphous materials have not been explored, however. Kwon *et al.* showed that a generative diffusion model can be used to generate amorphous carbon matching conditioned on characterization data.^[17] Nevertheless, the resulting structures still showed outlier environments, and it remained unclear whether generative models can be used to generate and reproduce the properties of more complex amorphous materials such as multi-component ones, especially when conditioned to processing parameters.

In this work, we proposed a suite of tests to validate the quality of generated amorphous structures. As a first figure of merit, structures are evaluated according to their short- and medium-range order, which includes pair distribution functions, bond angles, and network topology. These distributions are commonly used to understand and analyze the structure of amorphous materials, and often correlated with experimental observations. As a second tier of tests, we computed the macroscopic properties of simulated and generated structures such as their elastic tensor and plastic behavior. These mechanical properties are highly sensitive to outlier environments, as minor unphysical features can produce large variations in energy (see Appendix A.7 below), thus providing an excellent test case to assess the robustness of the generative models. Finally, we introduced an information-theoretical metric to evaluate whether structures are sampled from similar probability distribution of atomistic environments, which is useful to quantify novelty (see Appendix A.8 below). These strategies are summarized in Fig. 1d of the main text.

A.2 Importance of an extra noise in denoising process

Adding a high level of extra noise during the denoising process is critical to generate valid and novel amorphous structures. The noise magnitude σ is analogous to a temperature in conventional melt-quench molecular dynamics simulations, a connection that has been explored before in the context of Langevin dynamics within diffusion models.^[3] High σ values in the beginning of the denoising process are analogous to elevated temperatures, where atoms possess sufficient kinetic energy to escape local minima and break structural correlations imposed by the initial state. Furthermore, the progressive reduction of σ to zero during denoising is similar to the controlled cooling process in MD quenching. The only difference in this analogy is that the generative model can produce structures with different cooling rates at a fixed number of steps, whereas an MD-based melt-quench process requires long simulations to properly cool the structures. Denoising without extra noise resembles instantaneous quenching from a liquid to a glassy state, preventing adequate structural relaxation.

Partial pair distribution function (PDF) analysis in Fig. 15 shows that amorphous structures with superior structural quality are obtained when extra noise is employed during the denoising process. Importantly, when approximating the distribution of atomistic environments in the training data, the generative model makes no consideration about the energy of the system. Sometimes, this leads to a few outlier environments that, while rare, can be nonphysical for the simulated system, such as

Si-O bonds shorter than 1.5 Å in a-SiO₂ (Fig. 15a,b). While both approaches generate amorphous structures with reasonable short-range order, structures generated with extra noise exhibit denser bond pair distributions and more accurate peak intensities compared to reference simulations (Fig. 15c,d). Directly applying MD fine tuning process on structures generated without extra noise can eliminate unphysical short Si-O bonds and enhance partial PDF performance; however, the second coordination shell remains poorly reproduced (Fig. 15b). On the other hand, the small amounts of outlier environments in structures generated with extra noise can be readily corrected through picosecond-long MD simulations after the denoising process and show excellent agreement with the simulated reference (Fig. 15d).

As discussed in the main text and Methods, structural diversity of amorphous structures was quantified using overlap scores between six independently generated a-SiO₂ structures. Without extra noise, overlap scores were distributed either excessively high (approaching 100%, indicating mode collapse) or unusually low (below 45%, suggesting structural inaccuracy) as shown in Fig. 18a. Conversely, structures generated with extra noise exhibited consistent overlap scores around 70%, matching nearly perfectly the balance between diversity and validity observed in simulated structures (Fig. 18b). When compared to MD simulated reference structures, configurations generated without extra noise showed poor correlation (overlap scores around 29%, Fig. 19a), whereas those with extra noise maintained consistent accuracy with overlap scores around 70% (Fig. 19b). Altogether, these figures of merit (PDF analysis, overlap score distributions, and structural diversity metrics) provide evidence on the importance of using a large extra noise when generating amorphous materials using denoiser models.

A.3 Molecular dynamics simulations

All molecular dynamics (MD) simulations were performed using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software [45] (v. 17 Apr 2024). A fixed integration timestep of 1.0 fs was used throughout all simulations.

SiO₂ simulations: To generate training data for the SiO₂ systems, we conducted MD simulations of amorphous silica comprised of 3000 atoms across multiple cooling rates. We executed six independent simulations at each cooling rate, maintaining the same settings while varying initial configurations. We adopt the Jakse interatomic potential, [46] parametrized from *ab initio* calculations, to model a-SiO₂. Short-range interactions are calculated with an 8.0 Å cutoff. Coulomb interactions are calculated by adopting the Fennell damped shifted force model with a damping parameter of 0.25 Å⁻¹ and a global cutoff of 8.0 Å. Periodic boundary conditions were applied in all three directions during MD simulations.

All a-SiO₂ samples were prepared via conventional melt-quench method. Initial structures were generated by randomly placing SiO₂ molecules in cubic boxes using PACKMOL (v. 20.13.0), [47] enforcing a 2.0 Å minimum separation between molecules to prevent unphysical overlaps. Following energy minimization, configurations underwent sequential 50 ps relaxations in the canonical (*NVT*) and isothermal-isobaric (*NPT*) ensembles at 300 K. Complete melting was achieved with a 50 ps-long *NPT* equilibration at 5000 K and zero pressure, eliminating the “memory” of the initial configuration. Subsequently, liquid systems were cooled from 5000 to 300 K under zero pressure *NPT* conditions at rates of 10⁻¹, 10⁰, 10¹, and 10² K/ps. The obtained glass structures are further relaxed one last time at 300 K for 100 ps in the *NPT* ensemble. This process follows the standards from previous work [29] shown to reproduce multiple experimental features from a-SiO₂ glasses.

CuZr simulations: We adopted the published dataset from Wang *et al.* [42], which used a set of optimized embedded-atom method (EAM) potentials to simulate CuZr metallic glass models. [48] Specifically, we used two of their Cu₅₀Zr₅₀ configurations, each containing 5,000 atoms, to train our denoiser model. The potential from Cheng *et al.* [48] was later used to perform the short MD refinement and the fracture test (described below) for generated CuZr structures. The post-denoising MD simulation was 25 ps-long at the *NVT* ensemble, followed by a 25 ps-long simulation at the *NPT* ensemble, both at 300 K.

Computation of elastic properties: the stiffness tensor C_{ij} of the equilibrated glasses is computed by performing a series of 6 deformations (i.e., 3 axial and 3 shear deformations along the 3 axes) and calculating the curvature of the potential energy U :

$$C_{ij} = \frac{1}{V} \frac{\partial^2 U}{\partial e_i \partial e_j}, \quad (6)$$

where V is the glass volume, e is the strain, and i, j are the indices corresponding to each Cartesian direction. All simulated configurations exhibited near-complete isotropy. Bulk modulus (K), shear modulus (G), Young’s modulus (E), and Poisson’s ratio (ν) are then calculated from the stiffness tensors.

Fracture tests: fracture simulations are performed on different starting bulk configurations by deforming the structures along a single direction during an MD simulation. Uniaxial deformation is achieved by imposing a constant strain rate of 10^9 s^{-1} along the z direction while allowing lateral dimensions (x and y) to relax freely under zero lateral pressure, approximating realistic uniaxial tension conditions. The deformation process is controlled using an NPT ensemble maintained at 300 K and 1 bar. True stress is calculated from the negative z -component of the stress tensor.

A.4 Computational cost estimates

We used a single CPU core of Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz from UCLA’s Hoffman2 Cluster to estimate the computational cost of MD simulation and denoiser generation for 3000 atoms a-SiO₂. The entire MD simulation process includes initialization, melting, cooling and relaxation parts as detailed above, where only the cooling part is affected by the choice of cooling rate. The denoising process (inference) consists of a total of 3000 denoising steps (2900 steps with extra noise and 100 steps without extra noise). The trained denoiser model is also saved, loaded, and evaluated on the same single CPU for this comparison. In practice, we used GPUs for the generative inference process, which perform faster than using CPUs for inference of the GNNs.

To estimate the computational cost of traditional melt-quench MD simulations for large systems at slow cooling rates, we benchmarked structures at target sizes for 25 ps. All simulations include fixed computational overhead include 50 ps NVT initialization, 50 ps NPT initialization, 50 ps NPT melting, and 100 ps NPT relaxation (250 ps total). The quenching process duration varies with cooling rate, requiring 47, 470, 4,700, 47,000 and 470,000 ps for 10^{-2} , 10^{-1} , 10^0 , 10^1 , and 10^2 K/ps, respectively. Total computational cost is calculated by multiplying the average time per step obtained for a 25 ps MD simulation by the total number of steps to perform the simulation. For the large system estimate, the cost of performing a 25 ps MD simulation of 112,848 atoms within a-SiO₂ and the simulation settings described above is 4.77 CPU-h. This results on the total estimated cost of 89,702 CPU-h for the same size under 0.01 K/ps cooling rate described in the main text.

A.5 Structural analysis

All structural features of simulated a-SiO₂ structures were performed with six independent samples of 3000 atoms each. Samples are obtained either through the traditional melt-quench process, or using the generative models. All structural properties reported in the main text (e.g., Fig. 2[3]) are reported as either averages over these six samples (e.g., structural features in Fig. 2a-c) or distributions (e.g., mechanical properties in Fig. 2d). This allows us to perform statistically relevant comparisons between generated and simulated structures by accounting for the distribution of structures and properties in amorphous materials.

Pair distribution functions: short-range structural ordering was characterized through pair distribution function (PDF) analysis. All PDFs were calculated using OVITO[49] with 100 bins between 0 and 8 Å.

Ring size distribution: the medium-range order structure of amorphous silica was characterized through ring size distribution analysis. These calculations were performed using the RINGS package,[50] with Si-O bonds defined using a 2.0 Å cutoff distance for ring identification within the network structure, which is in agreement with the partial PDF for Si-O pairs in Fig. 2a.

Concentration of non-bridging oxygens: we quantified the concentration of non-bridging oxygens (NBOs) on generated mesoporous silica surfaces by calculating the ratio of NBO atoms to the estimated surface area of the pore. As described in the main text, the pore morphology exhibits density-dependent shape and distortions from ideal cylindrical structures, as illustrated in Fig. 13. To address this non-ideal geometry, we analyzed the radial distribution of Si and O atoms relative

to the central simulation axis. The effective pore radius was estimated as the distance at which the density of Si and O atoms plateaus, indicating the boundary between pore space and silica framework. Whereas our generation has been performed only for the SiO_2 composition, NBOs are of high relevance to applications of mesoporous silica, where they can become silanol groups in the presence of hydrogen.

Atomic volume: We employed Voronoi analysis in OVITO[49] to compute the atomic volumes of CuZr structures. Particle radii of 1.35 Å for Cu and 1.55 Å for Zr atoms are applied for analysis, with relative face area threshold set to 1%.

Voronoi indices: Voronoi indices for CuZr structures were computed using OVITO[49] for poly-disperse Voronoi tessellation with particle radii of 1.35 Å for Cu and 1.55 Å for Zr atoms. Surface atoms occupying less than 1% of the total surface area were excluded to minimize effect from experimental measurement and structural reconstruction errors. Voronoi indices are represented using reduced Schläfli notation, $\langle n_3, n_4, n_5, n_6 \rangle$, where n_i denotes the number of polyhedral faces containing i edges. The polyhedral face distribution quantifies the frequency of faces with varying edge numbers.

A.6 Information entropy and QUESTS method

Representation: the representation of atomic environments was computed as described in our previous work.[27] In summary, a number of $k = 32$ neighbors was used to represent the atomic environment, with a cutoff of 5 Å. No distinction was made between element types, which is implicitly recovered based on the coordination environments, and thus captured in the information entropy of the system.

Information entropy: the information entropy of descriptor distributions was computed as described before.[27] Given a set of feature vectors $\{\mathbf{X}\}$, their information entropy is computed as follows:

$$\mathcal{H}(\{\mathbf{X}\}) = -\frac{1}{n} \sum_{i=1}^n \log \left[\frac{1}{n} \sum_{j=1}^n K_h(\mathbf{X}_i, \mathbf{X}_j) \right], \quad (7)$$

where our choice of the kernel K_h is the Gaussian kernel,

$$K_h(\mathbf{X}_i, \mathbf{X}_j) = \exp \left(\frac{-\|\mathbf{X}_i - \mathbf{X}_j\|^2}{2h^2} \right), \quad (8)$$

with a constant bandwidth $h = 0.015 \text{ Å}^{-1}$, as studied before.[27]

We define the differential entropy $\delta\mathcal{H}$ of a data point $\mathbf{Y}$ with respect to a reference dataset $\{\mathbf{X}\}$ as

$$\delta\mathcal{H}(\mathbf{Y}|\{\mathbf{X}\}) = -\log \left[\sum_{i=1}^n K_h(\mathbf{Y}, \mathbf{X}_i) \right]. \quad (9)$$

Overlap: From the definition of $\delta\mathcal{H}$, the overlap between two discrete distributions of atomic environments $\{\mathbf{Y}\}$ and $\{\mathbf{X}\}$ is the fraction of environments $\mathbf{Y}_i \in \{\mathbf{Y}\}$ with $\delta\mathcal{H}(\mathbf{Y}_i|\{\mathbf{X}\}) \leq 0$. Therefore, a zero overlap between two distributions indicates that the distributions have disjoint supports, whereas a 100% overlap indicates that the two distributions have the same support, even when the point-wise probabilities are different. Within the context of generative models described in the main text, a high overlap indicates that the generated structures follow the same distribution of atomic environments, while a low overlap suggests that the structures do not share the same environments.

Units: Throughout this work, the natural logarithm was used for the entropy in Eq. (7), which scales the information to natural units (nats).

A.7 Sensitivity of macroscopic properties to outliers

While structures generated with the denoising process without extra noise can capture peak positions and demonstrate reasonable agreement with simulated partial PDFs (Fig. 15a), they fail to reproduce the macroscopic properties of amorphous silica. Macroscopic properties such as mechanical properties and stress-strain behavior are quite sensitive to structural fidelity and are thus excellent tests to the quality of generated structures. Nonphysical atomic environments resulting from imperfect denoising can produce unrealistic bond lengths and angles that remain nearly undetected at short- and medium-range order (PDF, bond angle distributions, and ring size distributions such as Fig. 2 of the main text). However, even a single structurally invalid atom can result in major deviations in macroscopic properties, severely limiting the utility of generated structures for subsequent investigations.

Figure 16 compares stress-strain curves for structures generated under identical conditions using four methods: without extra noise, with extra noise, and each condition followed by a short MD equilibration (refinement). Despite very similar partial PDF performance across all samples (Fig. 15), their mechanical properties differ dramatically. Structures generated without extra noise fail to exhibit realistic fracture behavior and yield strength compared to simulated references even after a post-generation MD equilibration to reduce the occurrence of outliers. Adding substantial extra noise during the denoising stage produces much more reasonable stress-strain curves, with brittle fracture characteristics. Both non-refined approaches exhibit residual stress due to fixed density constraints during denoising. In contrast, structures generated with extra noise and MD refinement result in similar stress-strain curves given the expected variability of ductility across different samples, reproducing both yield strength and elastic modulus.

Quantitative analysis of the elastic properties further demonstrates this sensitivity. For all a-SiO₂ structures, we computed the average bulk modulus (K), shear modulus (G), Young’s modulus (E), and Poisson’s ratio (ν) across simulated and generated results. Table 1 summarizes these results. For the reference simulated structures, we obtain $K = 61.2$ GPa, $G = 37.1$ GPa, $E = 92.4$ GPa, and $\nu = 0.247$, compatible with the mechanical properties for this interatomic potential. 51 Generated structures without extra noise, despite exhibiting somewhat reasonable PDFs (Fig. 15), produce severely aberrant values: $K = 166.7$ GPa, $G = 2.9$ GPa, $E = 8.7$ GPa, and $\nu = 0.491$. The post-generation MD refinement corrects the elastic properties of generated structures, resulting in values of $K = 63.2$ GPa, $G = 37.5$ GPa, $E = 93.9$ GPa, and $\nu = 0.256$. Interestingly, although these values are close to simulated reference, the failure of exhibiting fracture behavior (Fig. 16) demonstrates that multiple validation approaches are essential for assessing generated structure fidelity, and underscores the critical importance of adding extra noise during the generation process. Adding extra noise during denoising significantly improves the properties of generated structures, but still exist small discrepancies, leading to $K = 67.0$ GPa, $G = 32.6$ GPa, $E = 83.6$ GPa, $\nu = 0.287$. Finally, after a short MD equilibration, structures generated with extra noise achieve excellent agreement with simulated results: $K = 60.6$ GPa, $G = 37.8$ GPa, $E = 93.8$ GPa, and $\nu = 0.241$. These findings underscore that mechanical properties provide a stringent test of structural fidelity, revealing deficiencies that remain hidden in conventional structural analyses.

A.8 Overlap as novelty metric of amorphous structures

In the main text, we introduced the use of information theoretical quantities to measure the “novelty” of newly sampled amorphous configurations. In generative models for molecules, for instance, novel data points are molecular graphs distinct from the training set, 52 53 and thus can be trivially computed by comparing the molecular graphs. In generative models for inorganic crystalline materials, novelty is defined by comparing the space groups, atomic basis, compositions, and other discrete quantities that uniquely define a crystal. 5 6 Whereas this is not as trivial as in the case of molecules and can lead to deviations in the assignment of “novelty” of a crystal, 54 it still allows for quantitative comparisons of novelty in crystal structure generation.

However, in contrast to crystalline structures or molecular graphs, novelty is challenging to define in the amorphous space. Even amorphous materials obtained from identical simulation conditions will not be identical unless they use exactly the same initial configuration and random seeds, leading to an ill-defined concept of “novelty” in the amorphous materials space. Thus, analyzing whether a generative model can approximate the data distribution and produce new samples requires measuring probabilities in the data domain. To do that, we quantified the similarity between amorphous

configurations using an information-theoretical method. Specifically, we computed a probability distribution over environments for a-SiO₂ structures using information theory,^[27] and evaluated the similarity between distributions using an information-theoretical overlap score. This comparison between distributions resembles the Fréchet Inception Distance in generative adversarial networks,^[28] which accounts for novelty in terms of distribution statistics rather than point-wise metrics. As discussed in the main text, we propose that the overlap of distributions of atomic environments can provide insights on the “novelty” of amorphous structures. These overlaps are computed per pair of structures, and thus can be depicted as a non-symmetrical matrix, as $P(A|B) \neq P(B|A)$. Figure 17 shows how these matrices look like when six a-SiO₂ structures are compared against other six structures. The diagonal of Figs. 17a,b indicate that the overlap of a structure with itself is 100%, as expected. Off-diagonal elements show that simulated a-SiO₂ samples (“Sim | Sim”) exhibited, on average, 75% overlap with each other despite deriving from the same processing condition, with narrow variability between the overlaps (Fig. 17b). This result shows that, for 3000-atom systems sampled in our simulations, about 75% of the environments are consistently sampled, and 25% environments are usually considered “new” even among deterministic simulations. Similarly, generated samples showed 70% overlap with each other (“Gen | Gen” in Fig. 2e). This overlap indicates that the generative model is capturing both the information relevant to describe an amorphous state and the approximate expected novelty in the amorphous states.

To test for mode collapse, we also computed the overlap between the distribution of generated environments and simulated ones (“Gen | Sim”), shown in Fig. 17c. If the model displayed a mode collapse and only learned to reproduce the training data, this overlap would be close to 100%. On the other hand, a model that consistently produces unphysical data points would exhibit very low overlap with the simulated structures even when the “Gen | Gen” overlap is high. Finally, a model that perfectly reproduces the distribution of simulated data should exhibit an overlap close to 75%. In our case, the overlap between the generated and simulated structures was approximately 70%, thus demonstrating that our generated a-SiO₂ samples maintain excellent structural fidelity while preserving appropriate configurational diversity. Figure 2e of the main text summarizes these findings by comparing the distribution of overlaps.

A.9 Information theoretical analysis of conditionally generated silica glasses

Given the known relationships between free volume, configurational entropy, and cooling rates in glasses,^[30, 55] we used our information entropy approach as a surrogate of configurational entropy for glassy materials.^[27] As explained in Section A.8, this method computes an information entropy of distributions of atomic environments and thus translates atomistic systems into probability distributions. Within glassy materials, even at constant density, input cooling rates should be able to steer the final generated structure to different states containing different distributions of atomistic environments, and thus different values of configurational entropy.^[56] In particular, rapid cooling produces high-energy states with incomplete structural relaxation, leading to glasses with more diverse local atomic motifs compared to slowly cooled counterparts, which is quantified by a higher information entropy. The relationship between information entropy averaged across six generated structures, cooling rate, and density of generated structures is shown in Figure 3a of the main text. The results confirm that higher information entropy is achieved for higher cooling rates, even at constant densities, validating the model’s ability to produce conditional generation of cooling rates. For generated a-SiO₂ samples at the right densities (Fig. 20), average information entropy values were 6.87, 6.70, 6.53, and 6.43 nats for cooling rates of 10², 10¹, 10⁰, and 10⁻¹ K/ps, respectively. These results are in excellent agreement with the average information entropy values of 7.01, 6.76, 6.58, and 6.43 nats for the reference simulated structures.

A.10 Effect of cooling rate on simulated and generated properties

MD simulations are widely employed to study glass structure and properties, yet their inherent limitation to short timescales imposes unrealistically fast cooling rates on simulated structures.^[29] This constraint makes it challenging to provide a one-to-one comparison with experimental glasses processed under typical laboratory conditions. In simulations of a-SiO₂, high cooling rates typically introduce spurious structural defects, though bulk structural characteristics remain relatively stable despite thermal history. Conversely, properties such as density and thermal expansion coefficients exhibit strong dependence on cooling rate. Since our denoising approach generates structures at fixed density (i.e., number of particles and volume), we developed a regression model that enables

us to specify the density as a function of cooling rate (Fig. 14). Within the range of cooling rates of interest, simulated amorphous silica densities are reasonably linear with the logarithm of the cooling rate, simplifying the determination of initial densities for denoising. Although the density of silica does not follow a linear trend across a wider range of cooling rates, 24 they are valid within our study and simplify the analysis of model generalization. The post-denoising MD equilibration employs both NVT and NPT ensembles, but the latter typically only increases the density by around 0.01 g/cm^3 .

Figure 20 illustrates how the information entropy changes at the final steps of conditional denoising across different cooling rates. Following the methodology described in the Methods section, entropy values remain approximately 8.0 nats before step 2200, reflecting the maximum possible entropy with a system with 3000 atoms ($\ln 3000 = 8.0$). 27 As the noise is progressively lowered due to the noise schedule, the information entropy of the system decreases and structures develop local ordering, with slower cooling rates leading to lower final entropy values, all consistent with simulated trends.

Validation of structures as a function of cooling rate employs mean Si–O–Si bond angles and mechanical properties, including Young’s modulus (Fig. 6), bulk modulus (Fig. 7), shear modulus (Fig. 8) and Poisson’s ratio (Fig. 9) across different cooling rates. Generated structures accurately capture Si–O–Si bond angle trends with respect to cooling rate, demonstrating excellent agreement in both interpolation regions (0.3, 0.5, 3, 5, 30, and 50 K/ps) and “generalization” domains, i.e., targeted cooling rates outside of the interval of $[10^{-1}, 10^2]$ K/ps. Mechanical properties, which exhibit high sensitivity to structural accuracy (see Appendix A.7 above), remain physically consistent across all cooling rates for generated samples. On the other hand, outliers, such as structures generated without extra noise (Fig. 15a), exhibit aberrant elastic properties (Figs 6–9), despite having similar partial PDFs compare to simulated samples. This confirms that our conditioned structure generation produces structurally valid amorphous configurations across a range of cooling rates.

A.11 Size and shape effects in fracture of amorphous silica

Despite known variability in simulated plastic properties such as strength, generated structures consistently lead to stress-strain curves with good approximation of the elastic regime and ultimate tensile strength of the simulated material. This is true even at large length scales, when cracks are strongly influenced by structural outliers. Moreover, larger generated structures exhibit an increasingly brittle behavior, with substantial reduction in ductility from structures with 3,000 and 30,000 atoms, consistent with past experimental results. 57 Figure 12 shows the stress-strain curves, averaged from six independent tests, for the fracture of generated cubes ($L_z/L_x = 1.0$) with system sizes ranging from 3,000 to 30,000 atoms. Given the larger aspect ratio of the generated samples compared to cubic structures in Fig. 12, brittle fracture is observed at lower strains in Fig. 4a. This agrees with previous studies stating that the length along the deformation axis is the most important factor in minimize boundary effects and provide adequate volume for realistic crack propagation. 57

Table 1: Average mechanical properties (bulk modulus K , shear modulus G , Young’s modulus E , and Poisson’s ratio ν) of a-SiO₂ structures for different structure generation methods. All values are in GPa except for ν (dimensionless). Errors correspond to the standard deviation of properties across six generated and simulated samples. Samples generated with extra noise and MD refinement exhibit the exact distribution of elastic properties of the simulated reference configurations.

Method	K (GPa)	G (GPa)	E (GPa)	ν
Generated (no extra noise)	166.7 ± 8.2	2.9 ± 0.3	8.7 ± 1.1	0.491 ± 0.001
Generated (no extra noise) + MD refinement	63.2 ± 2.1	37.5 ± 2.3	93.9 ± 4.5	0.256 ± 0.016
Generated (with extra noise)	67.0 ± 8.8	32.6 ± 4.3	83.6 ± 8.7	0.287 ± 0.042
Generated (with extra noise) + MD refinement	60.6 ± 3.1	37.8 ± 0.8	93.8 ± 1.4	0.241 ± 0.014
Simulated reference	61.2 ± 2.9	37.1 ± 3.0	92.4 ± 5.6	0.247 ± 0.025

B Figures

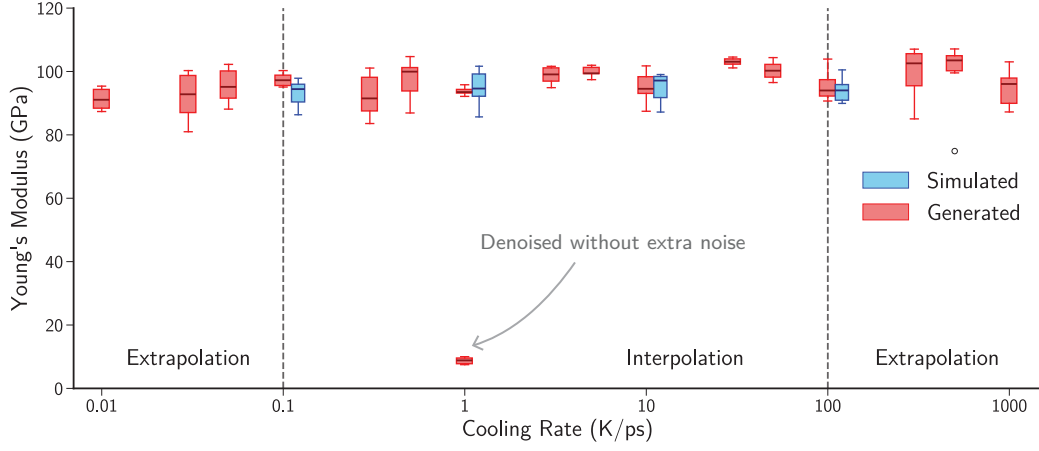


Figure 6: Young's modulus of simulated and generated a-SiO₂ samples across different cooling rates. The moduli are highly sensitive to outliers, yet the denoiser is able to generate structures with accurate Young's moduli compared to simulations, even outside the training domain (10^{-2} and 10^3 K/ps). The results of structures generated without adding extra noise demonstrate that high fidelity structures are necessary to reproduce reasonable elastic property.

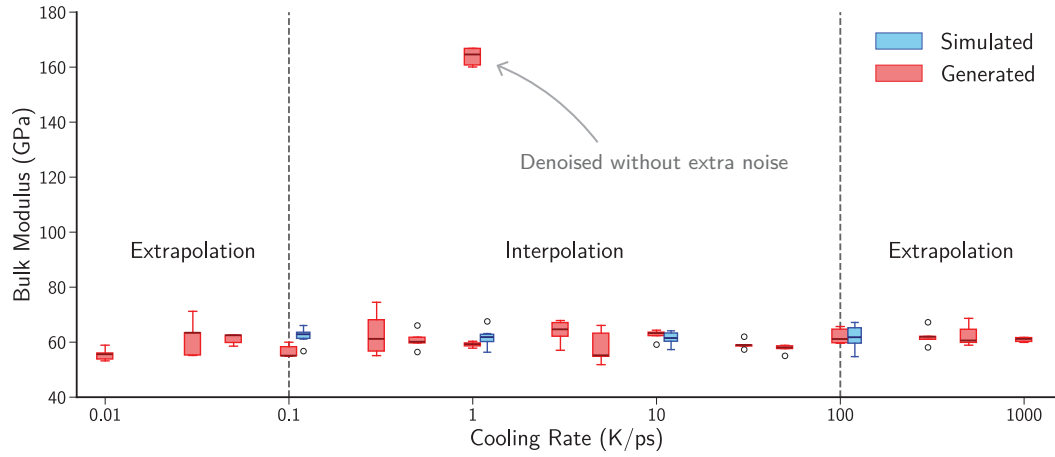


Figure 7: Bulk modulus of simulated and generated a-SiO₂ samples across different cooling rates. The moduli are highly sensitive to outliers, yet the denoiser is able to generate structures with accurate bulk moduli compared to simulations, even outside the training domain (10^{-2} and 10^3 K/ps). The results of structures generated without adding extra noise demonstrate that high fidelity structures are necessary to reproduce elastic properties.

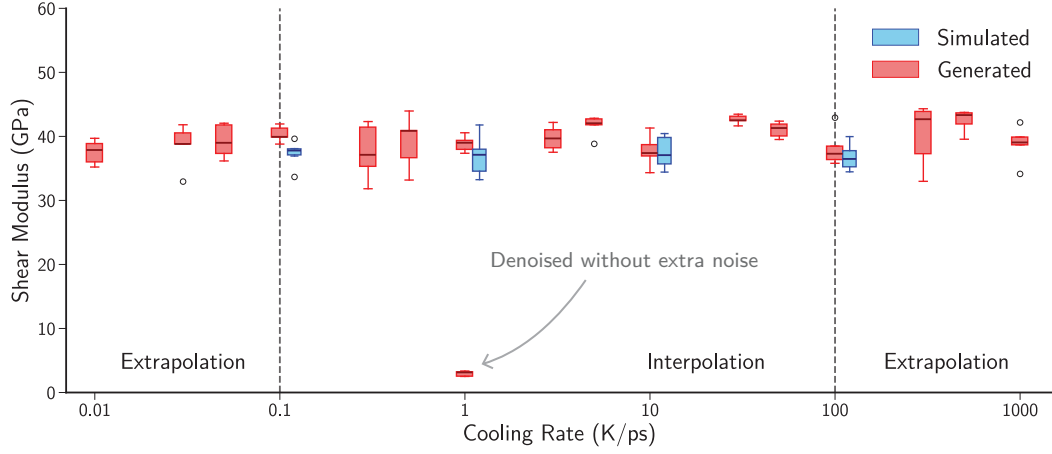


Figure 8: Shear modulus of simulated and generated a-SiO₂ samples across different cooling rates. The moduli are highly sensitive to outliers, yet the denoiser is able to generate structures with accurate shear moduli compared to simulations, even outside the training domain (10^{-2} and 10^3 K/ps). The results of structures generated without adding extra noise demonstrate that high fidelity structures are necessary to reproduce elastic properties.

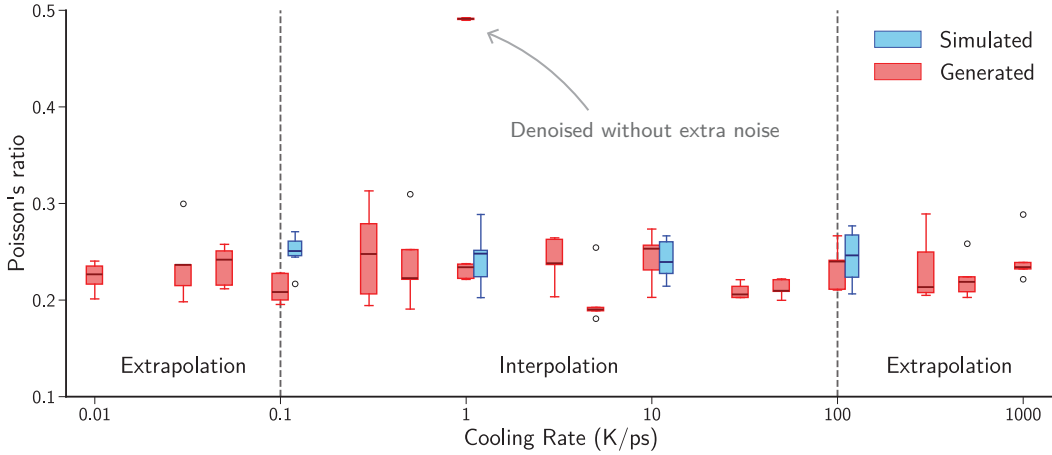


Figure 9: Poisson's ratio of simulated and generated a-SiO₂ samples across different cooling rates. The elastic properties are highly sensitive to outliers, yet the denoiser is able to generate structures with accurate Poisson's ratios compared to simulations, even outside the training domain (10^{-2} and 10^3 K/ps). The results of structures generated without adding extra noise demonstrate that high fidelity structures are necessary to reproduce elastic properties.

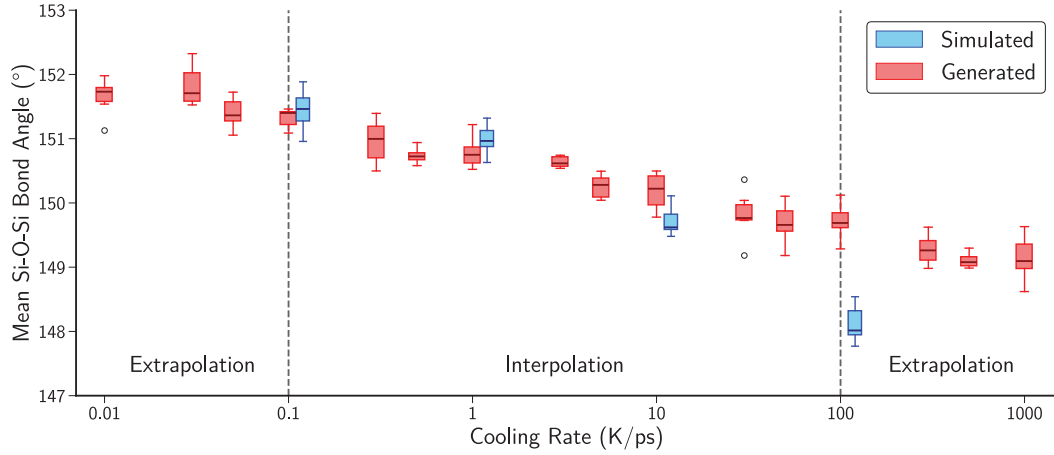


Figure 10: Mean Si–O–Si bond angle of a-SiO₂ structures as a function of cooling rate for simulated and generated samples. Each data point represents the average of six independent structures. The generated samples accurately reproduce the cooling rate dependence observed in simulated structures, with correct trends maintained even for cooling rates outside the training domain (10^{-2} and 10^3 K/ps).

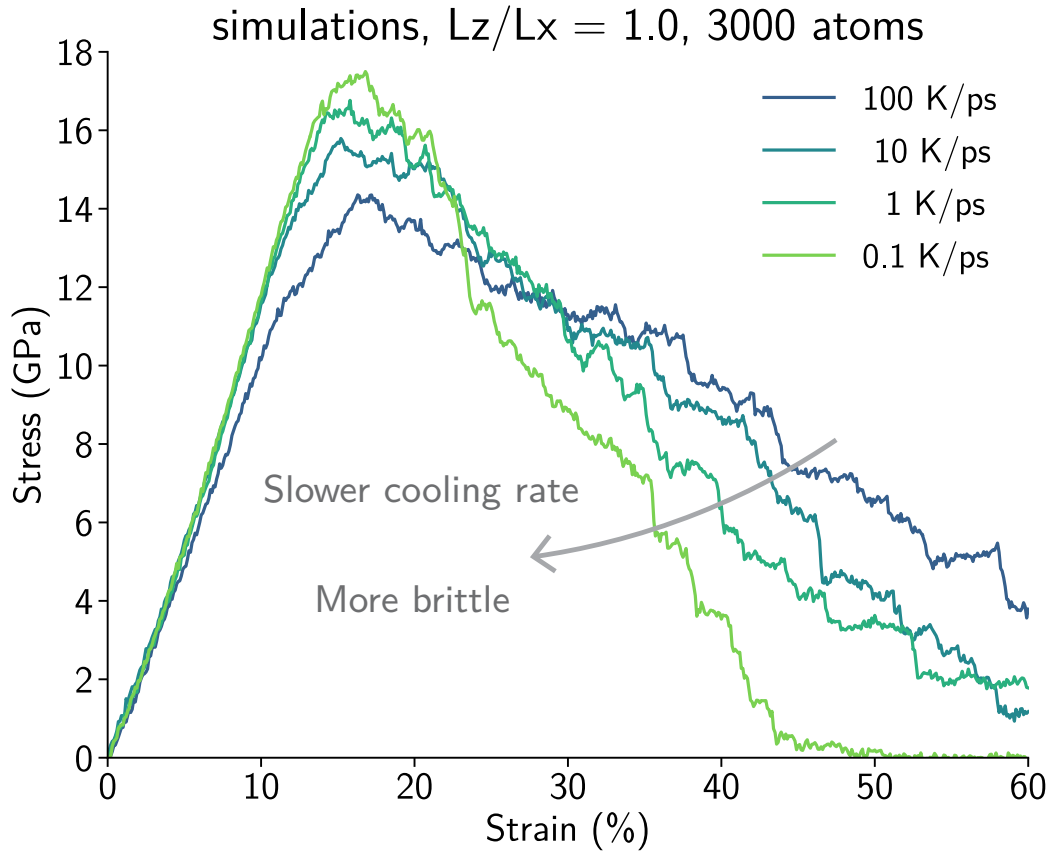


Figure 11: Stress-strain curves from simulated a-SiO₂ structures under different cooling rates show that slower cooling rate leads to a more brittle behavior. The initial structure was a cube with a total of 3000 atoms, and the fracture simulation was performed as described in the Methods. These results show that the ductility of the glassy system is high at this small scales, in contrast with results at large scales.

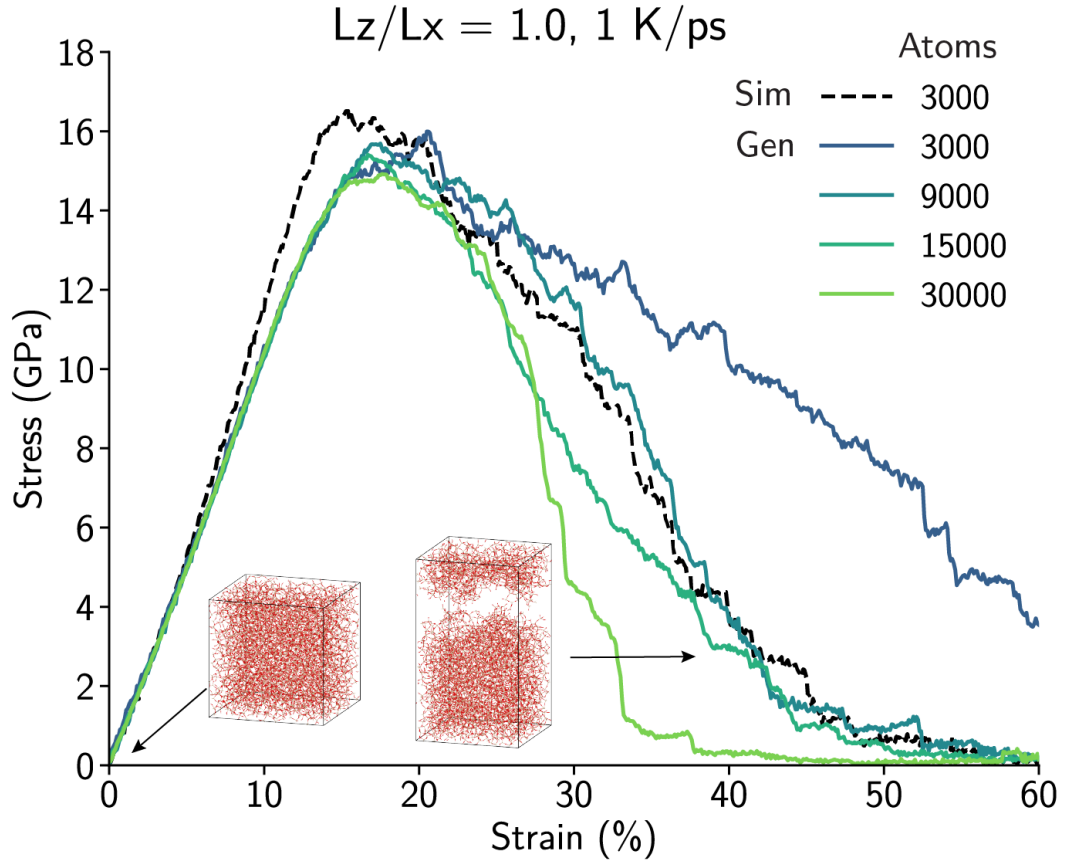


Figure 12: Stress-strain curves show that generated a-SiO₂ structures (solid lines) recover the expected trends of elastic and plastic deformation behavior across length scales, with more ductile fracture at smaller length scales. The stress-strain curve of a simulated a-SiO₂ structure with 3000 atoms is shown with a dashed line. The initial simulation box was a cube ($L_z/L_x = 1$).

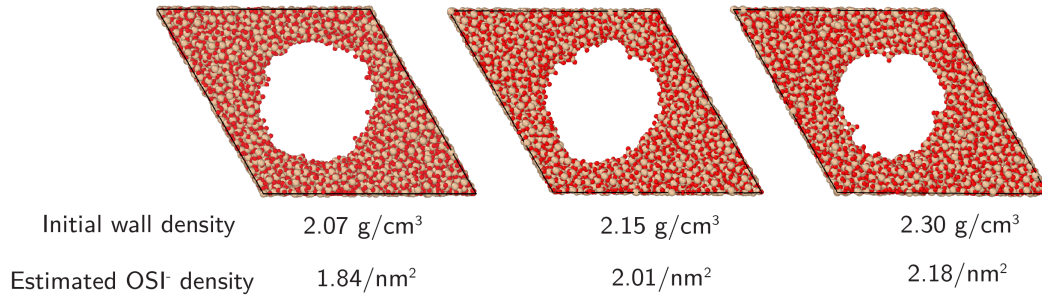


Figure 13: Mesoporous a-SiO₂ structures generated by the denoiser model at varying initial wall densities. All structures are generated from the same initial geometry and processing conditions. Higher wall densities result in correspondingly denser non-bonding oxygen groups on pore surface.

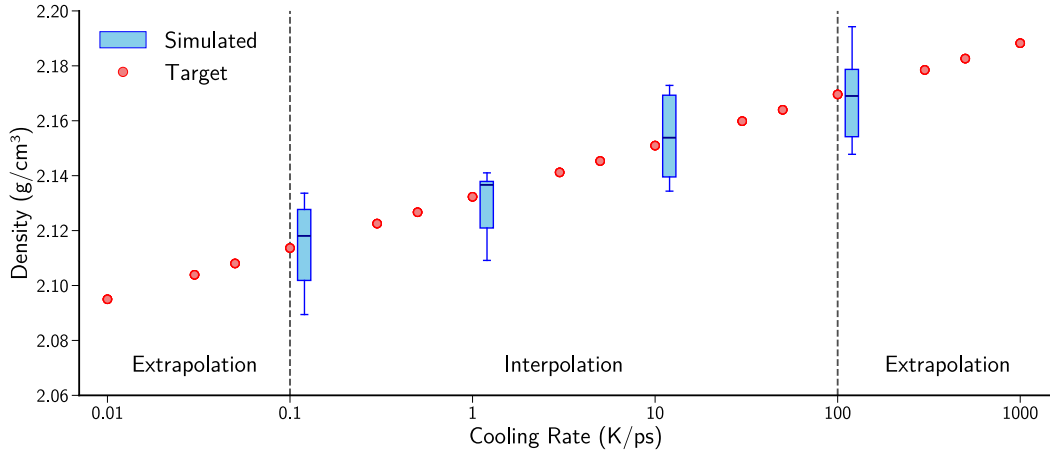


Figure 14: Densities of simulated a-SiO₂ structures and target densities across different cooling rates. Target densities are determined based on a linear regression model valid with simulated densities and extended to 0.01 and 1000 K/ps.

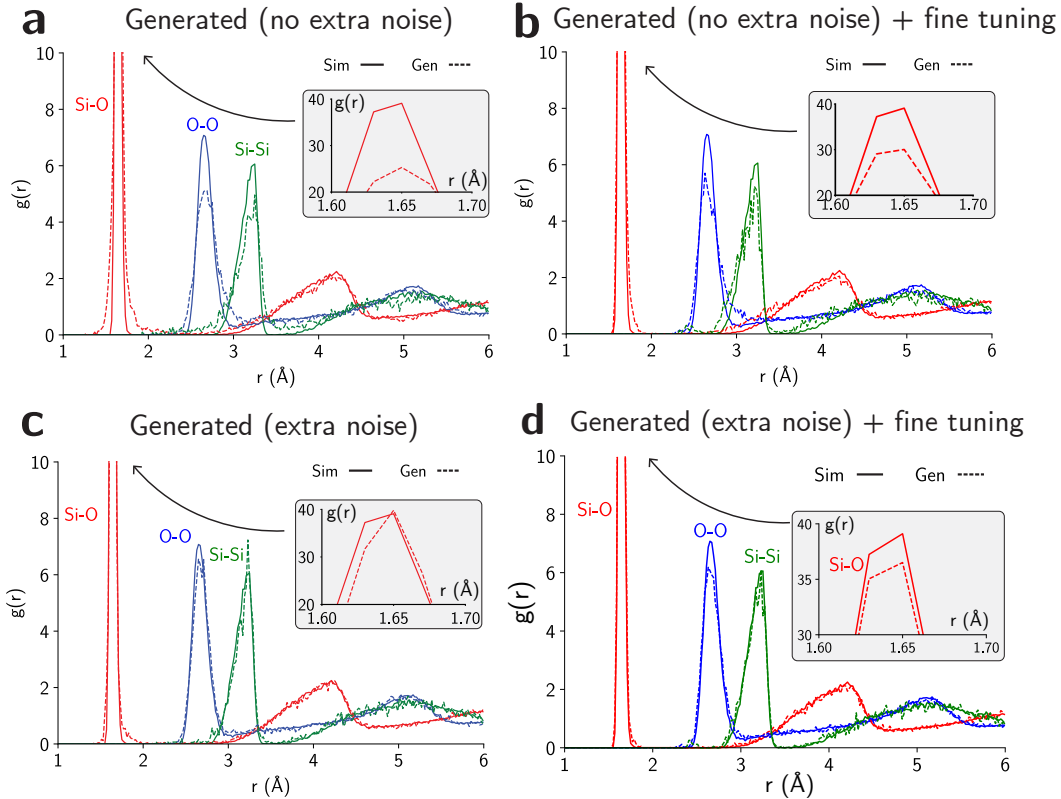


Figure 15: Comparison of partial pair distribution functions $g(r)$ between simulated and generated a-SiO₂ for cases: **a**, generated without adding extra noise, **b**, generated without adding extra noise, followed by MD refinement, **c**, generated with extra noise, and **d**, generated with extra noise and MD refinement). All cases demonstrate that generated configurations reproduce good structures according to structural quality metrics. Adding extra noise and fine tuning with MD removes most nonphysical behaviors (such as short Si–O pairs) and give accurate peaks for PDFs without intensity decreases.

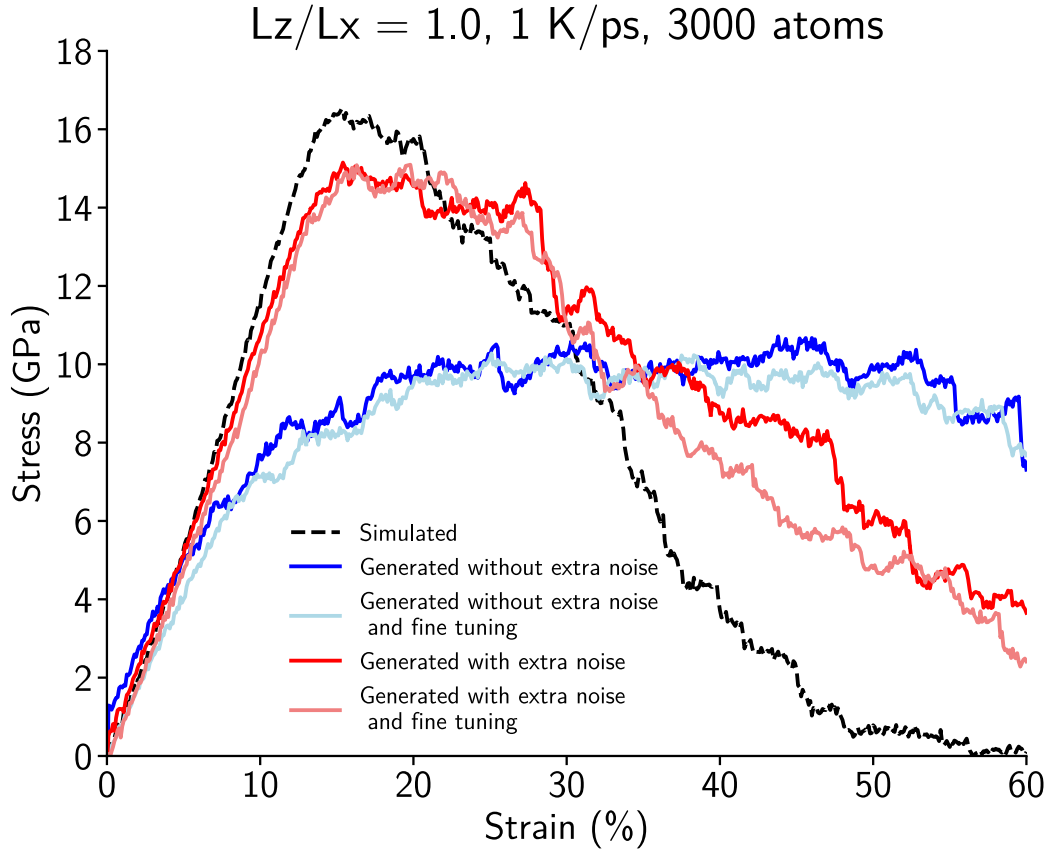


Figure 16: Stress-strain curves of simulated a-SiO₂ structure (dashed line) and generated structures (solid lines). Generated structures with extra noise and fine tuning show the best agreement compared to simulated results in terms of strength and ductility. Generated structures with extra noise show reasonable agreement compared to simulated result, but overestimate the stress at low strain and show more ductility. Generated structures without extra noise, even with fine tuning process, fail to reproduce the stress-strain behavior, showing a complete ductile behavior and underestimate the ultimate strength. All atomic structures were cube ($L_z/L_x = 1$) with 3000 atoms, and generated with a target cooling rate of 1 K/ps.

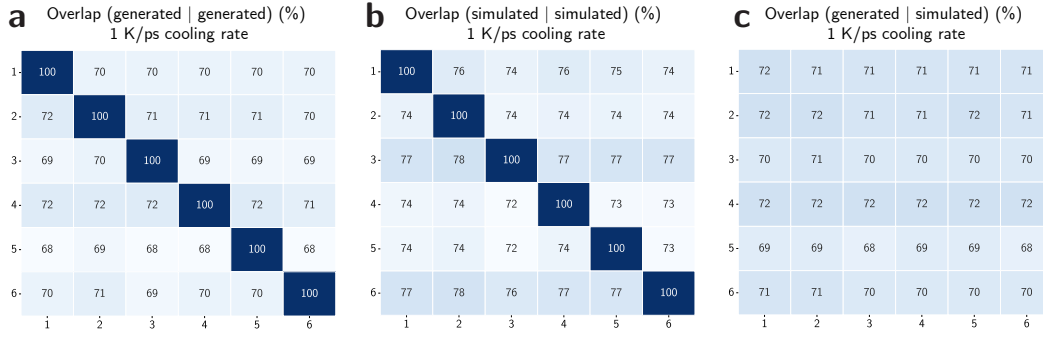


Figure 17: Overlap score matrices for **a**, generated | generated, **b**, simulated | simulated, and **c**, generated | simulated a-SiO₂ structures. Each matrix computes the overlap scores between two series of six independent simulations or generations. High overlap values indicate strong structural similarities. The results show two main results: (1) the distribution of generated atomic environments is very similar to the ones obtained from MD simulations, showing that the model produces valid amorphous structures; and (2) the generative model does not simply “copy-and-paste” the known amorphous structures, and instead produces novel configurations, with a novelty rate nearly identical to the ones obtained from simulations. All structures are prepared with 3000 atoms and under 1 K/ps cooling rate.

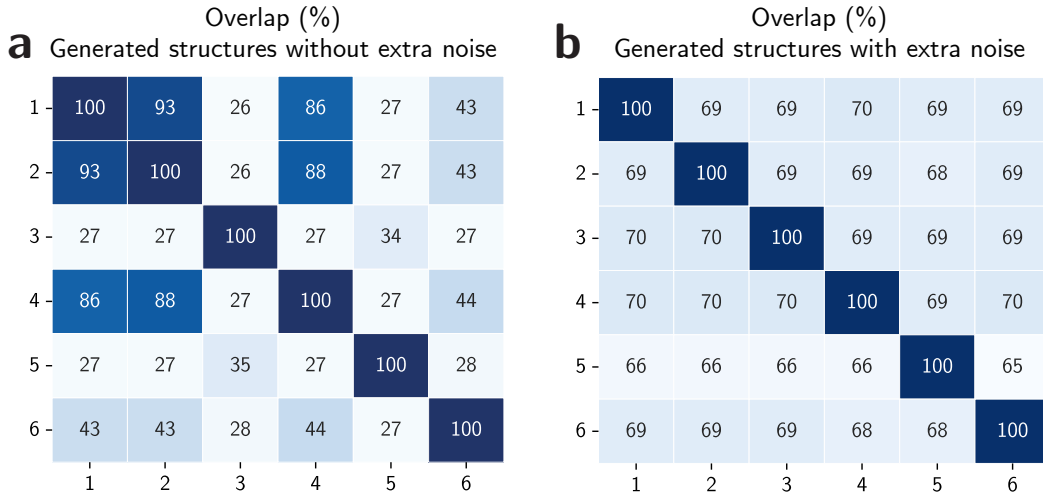


Figure 18: Overlap score matrices for generated a-SiO₂ structures **a**, without and **b**, with extra noise during the diffusion model. High overlap values indicate structural similarity, while low values indicate diversity. Without extra noise, structures show variable similarity. For instance, samples 1, 2, and 4 generated without noise are overly similar to each other, indicating a mode collapse into a single configuration. On the other hand, others differ substantially, indicating that the right distribution of environments is not captured. When extra noise is added, structures exhibit more consistent similarity across all samples.

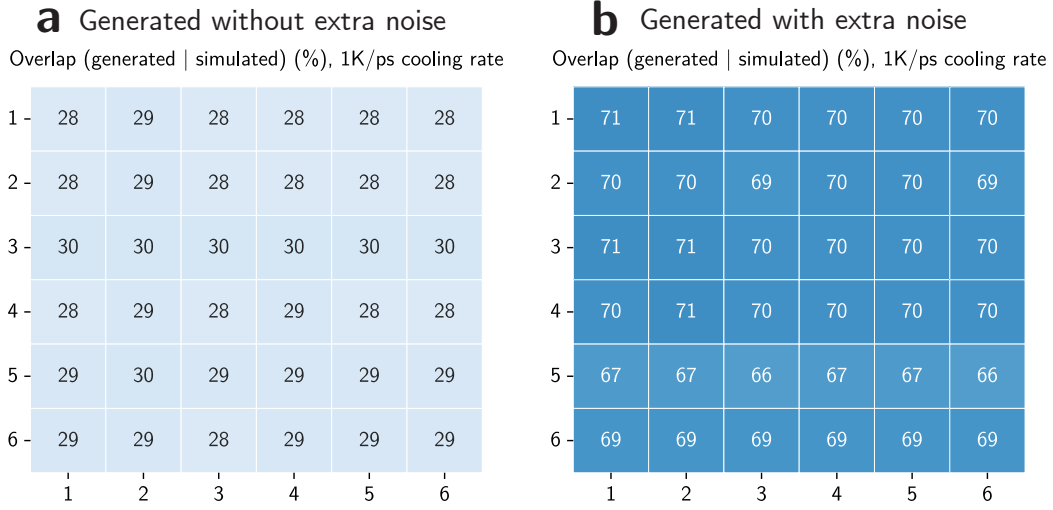


Figure 19: Overlap score matrices for generated against simulated a-SiO₂ structures **a**, without and **b**, with extra noise during generation. High overlap values indicate that the distribution of environments is similar. Generated samples without extra noise are mostly dissimilar to the simulated samples. On the other hand, the samples generated with extra noise are nearly as similar to simulated ones as simulated samples are to themselves.

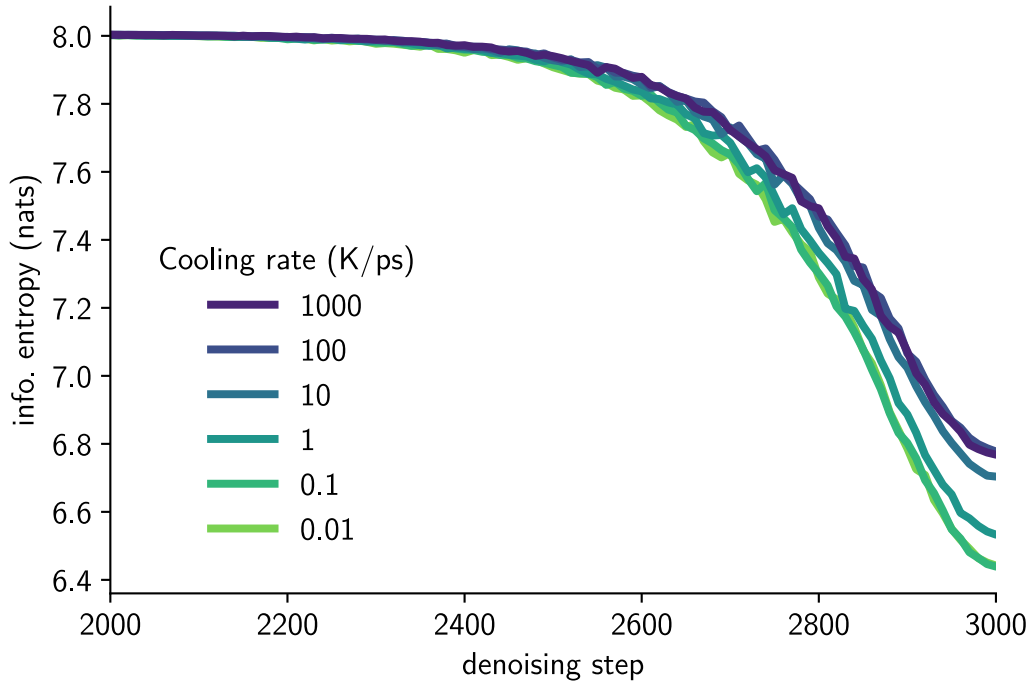


Figure 20: Information entropy of generated a-SiO₂ structures sampled along the denoising process with different target cooling rates. The denoising process was the same across the different conditions. The maximum information entropy that can be obtained for a system of this size with 3000 atoms is $\ln 3000 = 8.0$ nats.