

# Physical-regularized Hierarchical Generative Model for Metallic Glass Structural Generation and Energy Prediction

Qiyuan Chen<sup>1</sup>, Ajay Annamareddy<sup>1</sup>, Ying-Fei Li<sup>2</sup>, Dane Morgan<sup>1</sup>, Bu Wang<sup>1\*</sup>

<sup>1</sup>University of Wisconsin - Madison

<sup>2</sup>Stanford University

bu.wang@wisc.edu

## Abstract

Disordered materials such as glasses, unlike crystals, lack long-range atomic order and have no periodic unit cells, yielding a high-dimensional configuration space with widely varying properties. The complexity not only increases computational costs for atomistic simulations but also makes it difficult for generative AI models to deliver accurate property predictions and realistic structure generation. In this work, we introduce GLASSVAE, a hierarchical graph variational autoencoder that uses graph representations to learn compact, translation-, and permutation-invariant embeddings of atomic configurations. The resulting structured latent space not only enables efficient generation of novel, physically plausible structures but also supports exploration of the glass energy landscape. To enforce structural realism and physical fidelity, we augment GLASSVAE with two physics-informed regularizers: a radial distribution function (RDF) loss that captures characteristic short- and medium-range ordering and an energy regression loss that reflects the broad configurational energetics. Both theoretical analysis and experimental results highlight the critical impact of these regularizers. By encoding high-dimensional atomistic data into a compact latent vector and decoding it into structures with accurate energy predictions, GLASSVAE provides a fast, physics-aware path for modeling and designing disordered materials.

**Code** — <https://github.com/EricCH97/GlassVAE>

**Datasets** — <https://github.com/EricCH97/GlassVAE>

**Extended version** — <https://arxiv.org/abs/2505.09977>

## Introduction

Atomistic simulations of disordered materials—glasses in particular—are computationally intensive. The absence of long-range periodicity in these systems prevents the use of compact unit-cell descriptions, making new configurations have to be generated from first principles. (Bamer et al. 2023; Jung et al. 2025) While generative AI has been successfully applied to crystalline solids (Court et al. 2020; Li et al. 2025; Long et al. 2021; Noura, Sokolovska, and Crivello 2019a) and molecules by using structure representations like the Crystallographic Information File (CIF) (Flam-Shepherd

and Aspuru-Guzik 2023; Zeni et al. 2024) and Simplified Molecular Input Line Entry System (SMILES) (O’Boyle and Dalke 2018; Skinnider 2024; Weininger 1988; Sun et al. 2024), yet, no equivalent representation exists for inorganic glasses. Consequently, researchers rely on local geometric fingerprints—such as smooth overlap of atomic positions (SOAP) kernels (De et al. 2016; Lin et al. 2024; Bartók, Kondor, and Csányi 2013), or atom-centered symmetry functions (ACSF) (Mudassir et al. 2022)—which capture only short-range order and omit system-level geometry and related properties. Generative models trained solely on these local features often propose structures that appear locally plausible yet not satisfying global property constraints, hindering the discovery of new structures and materials.

In addition, glasses are inherently non-equilibrium systems with complex, high-dimensional configurational energy landscapes (Debenedetti and Stillinger 2001). Exploring these landscapes to identify realistic glassy structures is computationally intensive, as such structures occupy only a tiny subset of all possible configurations. Accurate generation of glass structures therefore requires the modeling to be aware of their configurational energy. Traditionally, molecular dynamics (MD) simulations have been the primary tool, where atomic interactions are explicitly simulated. By replicating the melt-quench process using MD, specific regions of the energy landscape are explored and yields structures resembling experimentally produced glasses. However, due to computational constraints, MD quenching rates are typically 6–9 orders of magnitude faster than the experimental rates (Binder et al. 2004). This results in structures with lower thermodynamic stability and altered properties (Yu et al. 2021). This timescale mismatch is also a major challenge for simulating key glass behaviors such as relaxation and aging (Yu et al. 2015; Berthier and Reichman 2023).

To address these challenges, graph-based representations combined with generative frameworks capable of predicting configurational energy offer a promising approach for modeling disordered systems like glasses. In these representations, atoms are treated as nodes and interatomic distances serve as invariant edge features. Through symmetric aggregation, such models inherently respect E(3) symmetries—translation, rotation, and permutation invariance for bulk atomic structures constrained by periodic boundary

\*corresponding author

conditions—enabling them to encode essential physical constraints directly into the architecture. (Satorras, Hoogeboom, and Welling 2022; Bao et al. 2023) This built-in invariance allows for the compression of high-dimensional atomic configurations into a compact, continuous latent space, significantly reducing computational overhead while retaining critical structural information. An explicit latent space is particularly useful for studying glass systems. Traversing this latent space enables the generation of new glassy structures and systematic exploration of their thermodynamic properties.

Despite the strong predictive performance of graph neural networks (GNNs) for atomic-scale property estimation (Li et al. 2024), embedding them within a generative modeling framework remains nontrivial. Hierarchical latent-space variational autoencoders (H-VAEs) (Klushyn et al. 2019; Vahdat and Kautz 2020; Ying et al. 2019) overcome this by learning multiscale latent representations: high-level variables capture global composition and medium to long-range features, while low-level variables encode local geometry. This architecture enhances expressiveness, mitigates posterior collapse, and produces physically plausible atomistic structures.

Here, we introduce GLASSVAE, a novel hierarchical variational autoencoder specifically designed to compress high-dimensional atomic graphs of disordered systems into a structured, low-dimensional latent space. GLASSVAE’s effectiveness stems from two key innovations: (i) **Physics-Informed Regularization**: A radial-distribution-function (RDF) loss ensures the preservation of essential short- and medium-range structural order (Rapaport 2004; Allen and Tildesley 2017), while an energy-regression loss aligns the latent embeddings with thermodynamic reality. (ii) **Specialized Hierarchical Latent Structure**: A graph-level latent variable captures global characteristics such as composition and the overall energy landscape, while a distinct edge-level latent variable refines local geometric details. This strategic division of responsibilities within the hierarchy effectively balances global energetic considerations with local structural fidelity, without unduly inflating the dimensionality of the latent space.

In this work, we demonstrate a prototype GLASSVAE on the CuZr metallic glass system. This system is selected to illustrate both the potential and the complexity of disordered alloys. (Pauly et al. 2010) These amorphous metals combine the high strength of crystalline metals with the formability of polymers. (Schroers et al. 2011) Their non-crystalline atomic arrangement underlies exceptional mechanical, thermal, and corrosion-resistance properties, driving their application in diverse fields including aerospace components, biomedical implants, and advanced sports equipment. (Trexler and Thadhani 2010; Bansal and Doremus 1986) Despite extensive study, the details of their structure-property relationship are still not well understood. (Starr et al. 2002; Cao et al. 2018) Through theoretical analysis and experiments on CuZr metallic glass, we demonstrate that GLASSVAE accurately reproduces potential-energy distributions, achieves low root-mean-square error (RMSE) in reconstructing atomic positions, and, crucially,

generates novel configurations that simultaneously adhere to RDF profiles and energetic constraints. By effectively unifying geometric and energetic priors within a single, tractable latent space, GLASSVAE establishes the first end-to-end generative framework specifically tailored for disordered inorganic systems. This work opens a new avenue for AI-guided discovery and design of complex glassy materials.

## Related Work

**Generative AI in Materials Science** GenAI models—including Generative Adversarial Networks (GANs) (Nouira, Sokolovska, and Crivello 2019b; Xu and Hu 2023), Variational Autoencoders (VAEs) (Gómez-Bombarelli et al. 2018; Luo et al. 2024), autoregressive models like Transformers (Wang et al. 2021), and diffusion models (Satorras, Hoogeboom, and Welling 2022; Bao et al. 2023)—have become powerful tools in materials design. These approaches are increasingly applied to sample novel compositions, predict structures, and estimate properties, often within a unified computational framework (Schleder et al. 2019). Significant breakthroughs have been demonstrated, particularly for crystalline materials, (Xie et al. 2019; Zeni et al. 2024; Ren et al. 2022) and for organic molecules, enabling de novo drug design and property optimization (Gómez-Bombarelli et al. 2018; Hoogeboom et al. 2022). Despite these successes, most generative efforts remain confined to systems exhibiting well-defined order, such as the periodic arrangements in crystals or the discrete bonding patterns in molecules. While generative techniques have been applied to glasses, (Li, Li, and Wang 2024; Cassar, Santos, and Zanotto 2021) these studies have largely focused on predicting macroscopic properties, generating microstructure morphologies, rather than atomic-level configurations (Jung et al. 2025).

**Graph Neural Networks** Graph neural networks (GNNs) have become indispensable tools in molecular and materials simulation, enabling accurate prediction of energies, forces, and other properties (Gilmer et al. 2017; Satorras, Hoogeboom, and Welling 2022; Sanchez-Gonzalez et al. 2020). Among all, Wang et al., combined GNNs with swap Monte Carlo to inversely design Cu–Zr metallic glasses with enhanced plastic resistance (Wang and Zhang 2021), and Li et al., applied GNNs to predict energy barriers and interpret the behavior of metallic glasses (Li et al. 2024). Yoshikawa et al., integrated self-attention into a GNN for structural classification of glass-forming liquids (Yoshikawa et al. 2025), while Bapst et al., trained a GNN solely on initial particle positions to predict long-time glass dynamics (Bapst et al. 2020). Although these studies demonstrate the power of GNNs for property prediction and classification in glassy systems, none simultaneously address structural generation. Our approach fills this gap by learning an internal generative mechanism that bridges state-of-the-art molecular-graph generative models with physically consistent atomic reconstructions.

**Physical-informed Descriptors** Global, physics-informed descriptors—such as Crystallographic Information Files (CIFs) for crystals (Flam-Shepherd and Aspuru-Guzik 2023; Zeni et al. 2024) and SMILES for

organic molecules (Weininger 1988; O’Boyle and Dalke 2018; Skinnider 2024)—require long-range order and thus fail for glasses. Instead, local geometric fingerprints like smooth overlap of atomic positions (SOAP) kernels (Bartók, Kondor, and Csányi 2013; De et al. 2016; Lin et al. 2024) or atom-centered symmetry functions (ACSF) (Mudassar et al. 2022) capture only short-range structure and omit system-level properties such as total potential energy. The radial distribution function (RDF) encodes short- and medium-range order in disordered materials. For example, Krykunov et.al., introduced atomic property-weighted RDFs with analytic Gaussian kernels that both outperform and accelerate Bag-of-Bonds for atomization energy prediction (Krykunov and Woo 2018), and Watanabe et.al., developed an on-the-fly RDF-based sampling method to flag anomalous configurations and build robust water potentials (Watanabe et al. 2024). To our knowledge, we are among the first to integrate an RDF-based loss into a variational graph model for glasses, combining RDF-driven regularization with graph representations to capture both local geometry and global energetics.

## Physical-regularized Hierarchical Graph Variational Autoencoder

In this work, we introduce GLASSVAE, a generative framework for disordered metallic-glass configurations that jointly captures local atomic geometry and global energetics. Our approach addresses three core tasks: (i) Reconstruction: Recover atom types, interatomic displacements and distances, and absolute positions; (ii) Energy prediction: Estimate the total potential energy of each configuration; (iii) Generation: Sample new, physically realistic atomic structures from the learned latent distribution. Formally, each molecular-dynamics snapshot is represented by  $(\mathbf{R}, \mathbf{t}, E)$ , where  $\mathbf{R} = \{\mathbf{r}_i \in \mathbb{R}^3\}_{i=1}^N$  (positions in a periodic box  $\mathbf{L} = [L_x, L_y, L_z]$ ),  $\mathbf{t}$  is the one-hot encoding of atom types, and  $E$  is the total potential energy. We learn  $\mathcal{E} : (\mathbf{R}, \mathbf{t}, E) \mapsto \mathbf{z}$ ,  $\mathcal{D} : \mathbf{z} \mapsto (\mathbf{R}, \mathbf{t})$ , and  $\mathcal{P} : \mathbf{z} \mapsto \mathbf{E}$  so that the latent variable  $\mathbf{z}$  retains all essential physical information for faithful reconstruction, accurate energy estimation, and the generation of novel, physically plausible structures.

### Graph Representation and Invariance

Physical interactions in materials depend only on atomic species and their relative arrangements, not on ordering, absolute position, or orientation. Therefore, atomistic representations and predictions must be invariant under permutation, translation, and rotation—swapping identical atoms, shifting all positions by  $\Delta \mathbf{t}$ , or applying any orthogonal rotation  $R$  leaves interatomic distances (and thus all physical properties) unchanged. To enforce these invariance, we convert each point cloud into an attributed graph  $\mathcal{G} = (V, E)$ , where Nodes  $i \in V$  carry one-hot features  $t_i$  encoding atomic species i.e.,  $x_i \in \{0, 1\}$ ; and Edges  $(i, j) \in E$  connect atoms within a cutoff distance (under periodic boundaries) and carry attributes

$$\mathbf{a}_{ij} = (\Delta \mathbf{r}_{ij}, d_{ij}), \quad \text{where } \Delta \mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j, \quad d_{ij} = \|\Delta \mathbf{r}_{ij}\|_2.$$

A graph neural network with symmetric (e.g., mean) message-passing layers aggregates these features, guaranteeing invariance to node ordering and rigid-body transformations. We further support these invariance claims with a formal theorem.

### Theorem 1 (Invariance of the Graph Representation)

Let an atomic configuration be given by positions  $\mathcal{R} = \{\mathbf{r}_i\}_{i=1}^N \subset \mathbb{R}^3$  and one-hot node features  $\{\mathbf{t}_i\}$ . Construct a graph  $G = (V, E)$  by connecting all pairs with  $\|\mathbf{r}_i - \mathbf{r}_j\|_2 \leq \text{CUTOFF}$ , and assign each edge the attribute

$$\mathbf{a}_{ij} = (\Delta \mathbf{r}_{ij}, \|\mathbf{r}_i - \mathbf{r}_j\|_2), \quad \text{where } \Delta \mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j.$$

Suppose the graph neural network updates and pools features using a symmetric (permutation-invariant) aggregation function  $S$  (e.g. mean or sum). Then the resulting representation is invariant to

**Permutation of nodes:** Since  $S$  does not depend on the ordering of its inputs, any re-indexing of the atoms leaves the pooled feature unchanged.

**Translation of the entire structure:** If  $\mathbf{r}'_i = \mathbf{r}_i + \mathbf{t}$  for all  $i$ , then

$$\Delta \mathbf{r}'_{ij} = (\mathbf{r}_i + \mathbf{t}) - (\mathbf{r}_j + \mathbf{t}) = \mathbf{r}_i - \mathbf{r}_j = \Delta \mathbf{r}_{ij},$$

so all edge-wise inputs to the network are identical.

**Rotation of the entire structure:** If  $\mathbf{r}'_i = R \mathbf{r}_i$  for some orthogonal  $R \in \mathbb{R}^{3 \times 3}$ , then

$$\Delta \mathbf{r}'_{ij} = R(\mathbf{r}_i - \mathbf{r}_j), \quad \|\Delta \mathbf{r}'_{ij}\|_2 = \|R(\mathbf{r}_i - \mathbf{r}_j)\|_2 = \|\Delta \mathbf{r}_{ij}\|_2,$$

so all distance-based edge features are preserved.

Theorem 1 thus demonstrates that our graph construction method and the use of symmetric aggregation functions yield a representation of 3D atomic configurations that is inherently invariant to permutations, translations, and rotations—properties that are essential for reliable modeling of physical systems.

### Model Architecture

To jointly capture the intricate local geometry and global properties (e.g., energies) of disordered metallic glass configurations, we propose a hierarchical graph variational autoencoder (GLASSVAE). Figure 1 sketches its architecture. Sec. states how we build a graph starting from an atomic configuration. From graph representations, we have

**Dual-path Encoder** The dual-path encoder that distills two complementary summaries. The graph-level encoding applies  $L$  layers of message passing to embed local neighbourhoods yielding a graph vector that two linear heads convert into the variational parameters  $\mu$  and  $\log \sigma^2$ . We could then reparameterize them into a decoder head  $z \sim \mu + \sigma$ . In parallel, the edge path feeds all  $a_{ij}$  through a residual stack of  $L_e$  blocks and averages the refined edge features into an edge descriptor  $s$ .

**Decoder** The decoder maps  $\mathbf{z}$  back to atomic species, i.e.,  $\hat{x} = D_1(z)$  and edge attributions, i.e.,  $\hat{a}_{ij} = D_2(z)$  thereby reconstructing the full configuration.

**Property Predictor** A separate fusion head consumes the concatenated vector  $\Phi(\mathbf{z} \parallel s)$  to predict the total

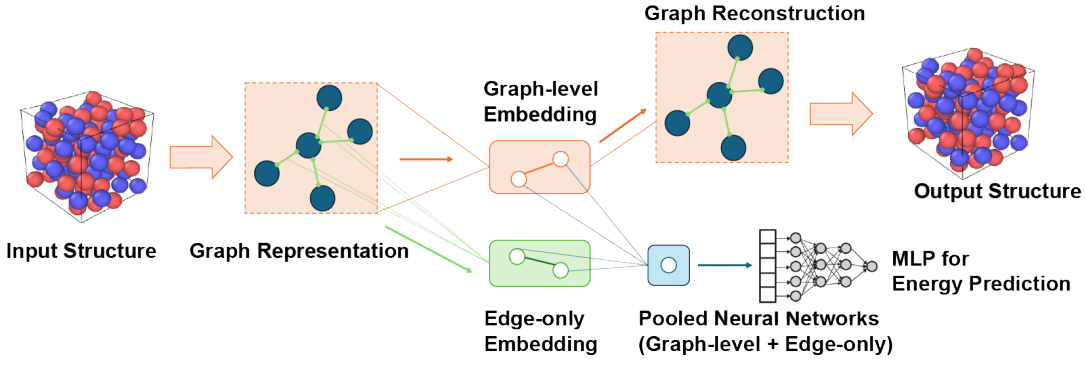


Figure 1: Schematic of Hierarchical Variational Autoencoder

energy  $\hat{E}$ , allowing the model to learn structure–property relationships directly from the latent space.

**Loss Function** The overall training objective is a weighted sum of several loss components:

$$\mathcal{L} = \alpha_{\text{node}} \mathcal{L}_{\text{node}} + \alpha_{\text{edge}} \mathcal{L}_{\text{edge}} + \alpha_{\text{energy}} \mathcal{L}_{\text{energy}} + \beta_{\text{KL}} D_{\text{KL}} + \alpha_{\text{RDF}} \mathcal{L}_{\text{RDF}}. \quad (1)$$

where each term enforces a different aspect of structural fidelity:

$\mathcal{L}_{\text{node}}$  is cross-entropy between ground-truth one-hot atom types and predicted class probabilities;  $\mathcal{L}_{\text{edge}}$  is MSE on predicted distances plus a cosine-similarity penalty to preserve directional consistency of edge features;  $\mathcal{L}_{\text{energy}}$  is MSE between predicted and true potential energies;  $D_{\text{KL}}$  is Kullback-Leibler (KL) divergence;  $\mathcal{L}_{\text{RDF}}$  is  $\ell_2$  distance between soft histograms of predicted and true pairwise distance distributions.

## Generation

Generation is one of the key features of GLASSVAE: its structured latent space enables both random and energy-conditioned synthesis of new, physically realistic configurations through the decoder. Here, we outline the two modes of operation. We also list the results of generation in Sec. .

**Random generation.** To produce novel structures, we draw

$$\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \text{ and set } \mathbf{z} = \boldsymbol{\mu}(x) + \boldsymbol{\sigma}(x) \odot \gamma \epsilon,$$

where “ $\odot$ ” denotes element-wise multiplication. This turns sampling into a differentiable operation with respect to  $\boldsymbol{\mu}$  and  $\boldsymbol{\sigma}$ . here  $\epsilon$  is a small perturbation. Then the sampled new structures could be decoded through  $\hat{x} = D_1(z)$  and  $\hat{a} = D_2(z)$ .

**Conditional generation.** To constrain the generated energy to  $[E_{\min}, E_{\max}]$ , we could use conditional generation. Formally, we assume a standard normal prior  $p(z) \sim \mathcal{N}(0, I)$ . To enforce  $\hat{E}(z) \in [E_{\min}, E_{\max}]$ , we initialize  $z$  at the encoder’s posterior mean and minimize

$$\mathcal{L}(z) = [\max(E_{\min} - \hat{E}(z), 0)]^2 + [\max(\hat{E}(z) - E_{\max}, 0)]^2 + \lambda \|z\|_2^2$$

for  $T$  steps. Decoding the optimized code  $z^*$  yields a structure whose energy lies within the target range.

## Theoretical Contribution of Physics-Informed Regularizer

Physics-informed loss terms (i.e. energy and RDF in our case) do more than improve reconstruction accuracy—they can associate the latent variables to physically admissible manifolds. Below we formalize a theoretical explanation: imposing an operator-based regularizer  $\|\mathcal{A}u - v\|_{L^2}$  induces a bound on the latent test error that decays with the training error and the quadrature resolution. In other words, every unit of training progress achieved by a physics-informed term translates directly into a proportionate tightening of the latent-space generalization gap.

**Functional setting.** Let  $\mathcal{U} = W^{s,q}(\mathbb{R}^3)$  and  $\mathcal{V} = L^p(\mathbb{R}^3)$  with  $1 \leq p, q \leq \infty$  and  $s \geq 0$ . Given a neural approximation  $\mathbf{u}(\cdot; \theta) \in \mathcal{U}$ , the operator  $\mathcal{A} : \mathcal{U} \rightarrow \mathcal{V}$  extracts a physical observable (e.g. predicted RDF, energy density), while  $v \in \mathcal{V}$  encodes the target dictated by first-principles or simulation data.

### Assumption 1 - Operator stability

There exists  $C_1 > 0$  such that  $\|\mathbf{u} - \mathbf{u}'\|_{\mathcal{U}} \leq C_1 \|\mathcal{A}\mathbf{u} - \mathcal{A}\mathbf{u}'\|_{\mathcal{V}}$  for all  $\mathbf{u}, \mathbf{u}' \in \mathcal{U}$ .

### Assumption 2 - Quadrature accuracy

A numerical quadrature with nodes  $\{\mathbf{x}_i, w_i\}_{i=1}^N$  satisfies  $|\int g - \sum_i w_i g(\mathbf{x}_i)| \leq C_2 N^{-\alpha}$  for all  $g \in \mathcal{V}$ , some  $C_2 > 0$  and  $\alpha > 0$ .

**Train–test decomposition.** Define the physics-informed training error

$$\mathcal{E}_{\text{train}}(\theta) := \sum_{i=1}^N w_i (\mathcal{A}\mathbf{u}(\mathbf{x}_i; \theta) - v(\mathbf{x}_i))^2,$$

and let  $\theta^*$  be a minimizer. The latent test error is  $\mathcal{E}_{\text{test}} := \|\mathbf{u}(\cdot; \theta) - \mathbf{u}(\cdot; \theta^*)\|_{\mathcal{U}}$ .

**Theorem 2 (Generalization bound)** *Under Assumptions –, we have*

$$\mathcal{E}_{\text{test}} \leq C_1 \mathcal{E}_{\text{train}}^{1/2} + C_1 C_2^{1/2} N^{-\alpha/2}.$$

Because both our energy- and RDF-based penalties satisfy the Lipschitz-stability condition (Assumption 1) and are

Method	Energy (eV/atom)		Paired Distances(Å)		Node
	RMSE	$R^2$	RMSE	$R^2$	BCE
Distance VAE (distance matrix as input)	11.85	0.89	0.24	0.77	—
GNN + MLP (no generation)	2.15	0.98	—	—	—
Single Graph-VAE (single latent space)	29.32	< 0	0.18	0.93	< 0.1
GraphVAE (no RDF)	0.37	> 0.99	0.14	0.94	< 0.1
<b>GlassVAE (best, + RDF)</b>	<b>0.32</b>	<b>&gt; 0.99</b>	<b>0.025</b>	<b>&gt; 0.99</b>	<b>0.091</b>

Table 1: Performance on the test set. We report RMSE (lower is better),  $R^2$  (higher is better), and Binary Cross Entropy (BCE) (lower the better).

evaluated in low-dimensional space ( $D = 3$ ), each physics-informed term directly drives down  $\mathcal{E}_{\text{train}}$ . By minimizing these terms we therefore *provably tighten* the upper bound on  $\mathcal{E}_{\text{test}}$ , improving generalization to unseen configurations. In this way, our physical regularizer are more than heuristics—they provide a principled mechanism for guaranteeing better predictive performance. This theoretical picture also has evidence from our experiments. Figure 2(c) visualizes the latent space of the full GLASSVAE, whereas Fig.S2 shows the same model *without* energy and RDF regularizes. The baseline VAE collapses into a tight cluster that fails to separate high- and low-energy structures. By contrast, the physics-aware GLASSVAE learns a more expressive, well-spread latent landscape that clearly distinguishes configurations by energy—exactly the discriminative structure predicted by the theorem.

## Experiments and Results

To evaluate the performance of GLASSVAE, we trained the model on molecular-dynamics (MD) trajectories of a  $\text{Cu}_{50}\text{Zr}_{50}$  metallic glass and then assessed it on an unseen test set. The system comprised 108 atoms with a glass-transition temperature  $T_g = 700$  K, with atomic interactions described using the embedded atom method (EAM) potential. The system was first equilibrated in the liquid state at 2000 K for 2 ns, then quenched to 1000 K in 50 K intervals at a cooling rate of 100 K per 6 ns and further cooled down in 20 K intervals at the rate of 100 K per 60 ns, all using the NPT ensemble. The simulations consisted of 108 atoms, and an MD time step of 1 fs is used. Periodic boundary conditions are employed in all three directions. We employed trajectories under 700 K, 760 K, 840 K, 920 K, 1000 K, with a total samples number of 50,000. In order to show the scalability of performance of GLASSVAE, we also evaluated GLASSVAE on two independent  $\text{Cu}_{50}\text{Zr}_{50}$  metallic-glass trajectories at 720K and 960K, both initiated from distinct starting configurations.

In summary, GLASSVAE’s hierarchical latent space combined with physics-informed regularization provides the flexibility needed for complex glassy systems and outperforms standard VAE variants (see Table.1), achieving superior accuracy in both property prediction and graph reconstruction. We show the detailed experimental results below and discuss them in more detail in the next section.

## Energy Prediction and Latent Space Exploration

Another core aim of GLASSVAE is accurate potential-energy prediction. On the test set (Figs. 2a–b), our model achieves an  $R^2 > 0.99$ , producing energy estimates whose overall distribution is virtually indistinguishable from the ground truth; even within defined subgroups (Fig.S1), it maintains  $R^2 \approx 0.85$ . Overall, we achieved a  $RMSE = 0.32$  among test dataset, which means on average a  $\sim 0.025$  eV reconstruction error on average. A UMAP projection of the latent codes (Fig. 2c), colored by true energy, reveals two well-separated clusters corresponding to configurations below and above the glass-transition temperature (McInnes, Healy, and Melville 2020). In contrast, the VAE without physics-informed regularization (Fig.S2) collapses into a single undifferentiated cluster. These results demonstrate that jointly training on energy not only delivers state-of-the-art accuracy but also embeds meaningful energetic structure into the latent space, enabling clear separation of physical regimes in accordance with our theoretical findings (Sec. ).

## Graph Reconstruction

In order to explore the reconstruction accuracy from GLASSVAE, we primarily looked into the correct identification of atomic species and the accurate recovery of their pairwise  $L_2$  distances. Accordingly, we evaluate whether our model can faithfully reconstruct both node features (atomic types) and edge attributes (interatomic distances). Overall, we achieved a  $RMSE = 0.025$  among test dataset, which means on average a  $\sim 0.025$  Å reconstruction error on average. Fig. 3 shows an example structure in the testing dataset. The model achieves excellent fidelity: it recovers atomic species with over 95 % accuracy, it does a good job in predicting the distances with a over 0.99  $R^2$  and a rather small RMSE. Fig. 3 (b) also exhibits the reconstructed distances follows a close similar distributions as original input. These findings confirm that GLASSVAE preserves both the structural and physical information of the original graph throughout the decoding process.

## Sampling from the Latent Space

A key advantage of VAEs is their ability to generate novel samples. Here, we demonstrate these generative capability via both conditional and random sampling: **Random Sampling** Random generation involves sampling from a simple standard normal distribution and then transforming this

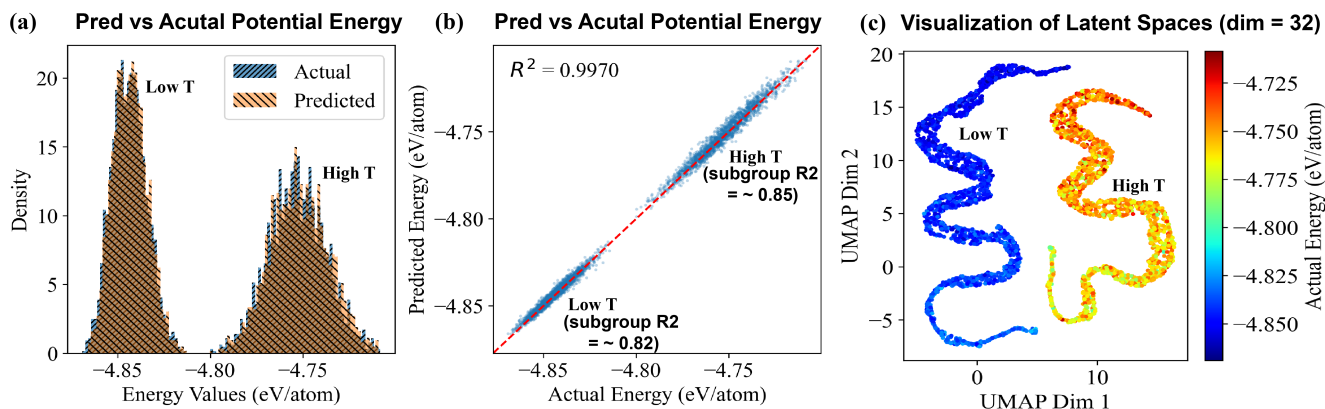


Figure 2: Energy Predictions and Latent Space Visualization. (a-b) Predicted vs actually potential energy; (c) visualization of graph latent spaces via UMAP projection

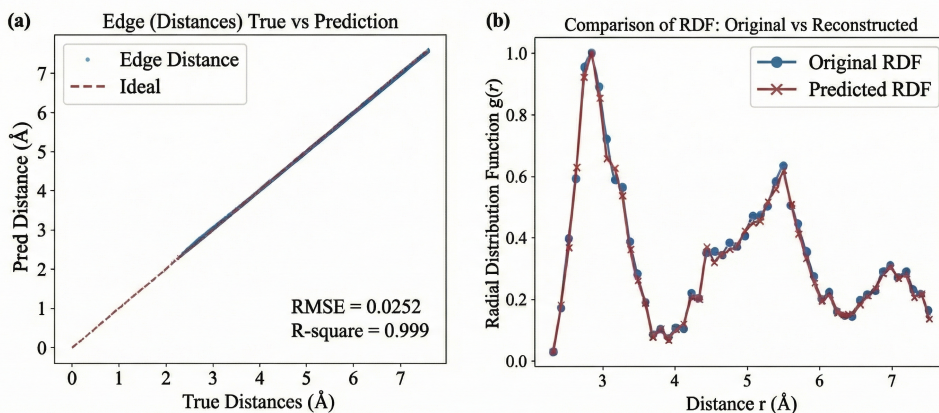


Figure 3: Edge Reconstruction Example (Distance unit Å). (a) shows the reconstructed distributions v.s. the ground truth. (b) Comparison of original RDF v.s. reconstructed

sample using learned mean and variance parameters, effectively drawing from the complex distribution learned by the model. Specifically, we sample directly from the standard normal distribution (the prior). We can also introduce controlled variations by adding a small amount of scaled noise to this initial sample before decoding. This perturbed sample is then passed through the decoder network to produce a final structure that represents a smooth deviation from the unperturbed version. Fig. 4 (a - b) show the results of 100 randomly generated samples and their calculated potential energy distributions. Fig. 4 (a) visualizes the new structures in latent space (cluster near known trajectories), indicating the generation of similar but novel structures. Moreover, the calculated potential energies of these generated structures (Fig. 4 (b)) align well with the energy distribution of the training data, suggesting they are physically meaningful.

**Conditional Generation** Instead, we could also generate atomic configurations whose potential energies lie within a specified interval by performing gradient-based refinement directly in the GLASSVAE latent space. Figure 4 (c) illustrates this process for a target energy of  $E = -4.87 \pm 0.01$  eV/atom. The five generated structures are marked in latent space as red 'x' symbols.

Both conditional generation and latent space optimization leverage the co-training of the energy prediction head to structure the latent space. By enforcing an energy constraint, as we discussed earlier in Sec. , the generated configurations are more likely to be physically realistic and adhere to the desired potential energy characteristics. This approach provides flexibility in generation—either directly conditioning on a target energy or refining latent samples post hoc via gradient-based optimization.

### Dimension Choice and Scalability Discussion

The choice of latent dimension is critical (Fig. S4): energy-prediction and edge-reconstruction errors decrease sharply as the latent dimension increases after  $dim = 10$ , before plateauing in the 16–32 range. A latent code in this latter range effectively captures most structural variability without superfluous model capacity. Furthermore, co-training the decoder with a differentiable energy head invests the latent manifold with thermodynamic significance. By using gradient-steering to guide latent codes into specific energy windows, GLASSVAE can synthesize configurations of predetermined stability. This allows for the rapid exploration of low-energy basins, offering an alternative to computation-

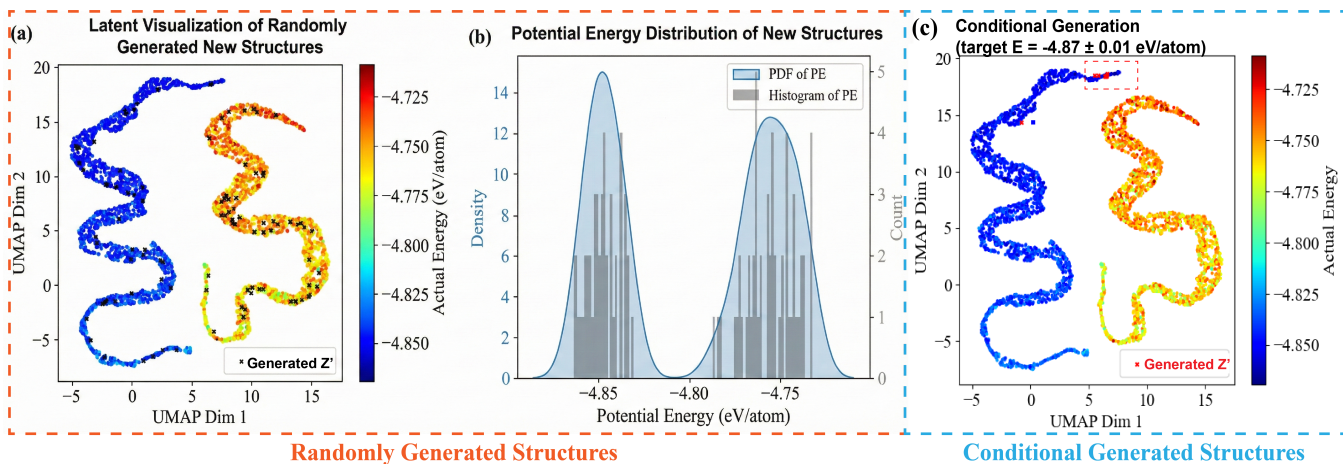


Figure 4: New Structure Generation. (a) Visualization of random sampled new structures in latent spaces (marked as black boxes); (b) Distribution of calculated potential energy of new structures, a kernel density estimate (KDE) was plot to show the continuous probability density (PDF) of new structures’ potential energies;(c) Conditional generation at  $E_{\text{tar}} = -4.87 \pm 0.01$  eV/atom. New structures are marked as red cross inside the labeled box

ally intensive molecular dynamics sampling.

Although randomly picked snapshots from a single trajectory are structural different, testing on held-out data alone may not fully demonstrate GLASSVAE’s scalability. To address this, we evaluated GLASSVAE on two independent  $\text{Cu}_{50}\text{Zr}_{50}$  metallic-glass trajectories at 720K and 960K, both initiated from distinct starting configurations. As shown in Fig.S3, the model’s energy predictions for these external trajectories still show accurate prediction patterns as original test set, confirming GLASSVAE’s ability to generalize to simulations with markedly different structural evolution. We observe a modest accuracy drop on the new testing trajectories, which likely reflects the limited structural diversity of our training set—only five trajectories at distinct temperatures. Due to computational constraints, we restricted our experiments to these five temperature trajectories to demonstrate crystals or SMILES for molecules. To bridge this gap, we introduce GLASSVAE’s capabilities. Incorporating additional trajectories and a wider range of conditions in future work should further enhance the model’s robustness.

## Conclusion and Discussion

Generative modeling of glassy materials has long been hindered by their disordered nature and the lack of long-range periodicity, which precludes compact encodings such as CIF for crystals or SMILES for molecules. To bridge this gap, we introduce GLASSVAE: a permutation-, translation-, and rotation-invariant graph variational autoencoder with hierarchical latent spaces. The multi-level embedding employed in this model is essential because a single latent space rarely offers sufficient flexibility to capture both the geometric complexity and property variability of glassy systems, making it challenging to achieve high-fidelity structure reconstruction and accurate energy prediction simultaneously (Table 1). We further bolster GLASSVAE with physics-informed losses—energy regression, radial distri-

bution function matching, and positional consistency—to ground its generative capabilities in physical reality.

Evaluated on a 108-atom  $\text{Cu}_{50}\text{Zr}_{50}$  dataset, GLASSVAE demonstrates high fidelity, reconstructing atomic identities and pairwise distances with an  $R^2 \sim 0.99$ . It also predicts potential energy with a Root Mean Square Error (RMSE) of 0.32 eV/atom for glass configurations over a wide thermodynamic stability. Beyond precise reconstructions, GLASSVAE facilitates both unconditional sampling and energy-conditioned refinement. This capability produces new atomic configurations whose energies and radial-distribution functions are consistent with simulation data, underscoring how physics-based regularizers significantly enrich and physically ground the latent space.

To our knowledge, GLASSVAE is the first architecture for inorganic glasses to unify structure generation and property prediction within a single, thermodynamically meaningful latent space. In addition, we also demonstrate GLASSVAE’s scalability to new  $\text{Cu}_{50}\text{Zr}_{50}$  trajectories initialized from different configurations. While current benchmarks use limited data from a 108-atom system to showcase the initial promise, future work can extend to larger, more complex glasses, disentangle latent factors (e.g., composition, local strain), incorporate additional thermodynamic observables (free-energy proxies, elastic constants), and integrate active-learning loops with simulations to further enhance fidelity and guide sampling in high-dimensional configuration space. Overall, we believe our approach paves the way for accelerated discovery of disordered materials by enabling the simultaneous mapping of energy landscapes and the proposal of candidate structures for further computational or experimental investigations.

## Acknowledgments

This research was primarily supported by NSF through the University of Wisconsin Materials Research Science and Engineering Center (DMR-2309000).

## References

- Allen, M. P.; and Tildesley, D. J. 2017. *Computer simulation of liquids*. Oxford university press.
- Bamer, F.; Ebrahim, F.; Markert, B.; and Stamm, B. 2023. Molecular Mechanics of Disordered Solids. *Arch Computat Methods Eng*, 30(3): 2105–2180.
- Bansal, N. P.; and Doremus, R. H. 1986. Chapter 1 - Introduction. In Bansal, N. P.; and Doremus, R. H., eds., *Handbook of Glass Properties*, 1–3. San Diego: Academic Press. ISBN 978-0-08-052376-7.
- Bao, F.; Zhao, M.; Hao, Z.; Li, P.; Li, C.; and Zhu, J. 2023. Equivariant Energy-Guided SDE for Inverse Molecular Design. ArXiv:2209.15408 [physics].
- Bapst, V.; Keck, T.; Grabska-Barwińska, A.; Donner, C.; Cubuk, E. D.; Schoenholz, S. S.; Obika, A.; Nelson, A. W. R.; Back, T.; Hassabis, D.; and Kohli, P. 2020. Unveiling the predictive power of static structure in glassy systems. *Nature Physics*, 16(4): 448–454. Publisher: Nature Publishing Group.
- Bartók, A. P.; Kondor, R.; and Csányi, G. 2013. On representing chemical environments. *Physical Review B*, 87(18): 184115. ArXiv:1209.3140 [physics].
- Berthier, L.; and Reichman, D. R. 2023. Modern computational studies of the glass transition. *Nature Reviews Physics*, 5(2): 102–116. Publisher: Nature Publishing Group.
- Binder, K.; Horbach, J.; Kob, W.; Paul, W.; and Varnik, F. 2004. Molecular dynamics simulations. *Journal of Physics: Condensed Matter*, 16(5): S429.
- Cao, Y.; Li, J.; Kou, B.; Xia, C.; Li, Z.; Chen, R.; Xie, H.; Xiao, T.; Kob, W.; Hong, L.; Zhang, J.; and Wang, Y. 2018. Structural and topological nature of plasticity in sheared granular materials. *Nature Communications*, 9(1): 2911. Publisher: Nature Publishing Group.
- Cassar, D. R.; Santos, G. G. d.; and Zanotto, E. D. 2021. Designing optical glasses by machine learning coupled with a genetic algorithm. *Ceramics International*, 47(8): 10555–10564. ArXiv:2008.09187 [cond-mat].
- Court, C. J.; Yildirim, B.; Jain, A.; and Cole, J. M. 2020. 3-D Inorganic Crystal Structure Generation and Property Prediction via Representation Learning. *J. Chem. Inf. Model.*, 60(10): 4518–4535.
- De, S.; Bartók, A. P.; Csányi, G.; and Ceriotti, M. 2016. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics*, 18(20): 13754–13769. ArXiv:1601.04077 [cond-mat].
- Debenedetti, P. G.; and Stillinger, F. H. 2001. Supercooled liquids and the glass transition. *Nature*, 410(6825): 259–267. Publisher: Nature Publishing Group.
- Flam-Shepherd, D.; and Aspuru-Guzik, A. 2023. Language Models Can Generate Molecules, Materials, and Protein Binding Sites Directly in Three Dimensions as XYZ, CIF, and PDB Files. ArXiv:2305.05708.
- Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; and Dahl, G. E. 2017. Neural Message Passing for Quantum Chemistry. ArXiv:1704.01212 [cs].
- Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; and Aspuru-Guzik, A. 2018. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science*, 4(2): 268–276. Publisher: American Chemical Society.
- Hoogeboom, E.; Satorras, V. G.; Vignac, C.; and Welling, M. 2022. Equivariant Diffusion for Molecule Generation in 3D. ArXiv:2203.17003 [cs].
- Jung, G.; Alkemade, R. M.; Bapst, V.; Coslovich, D.; Filion, L.; Landes, F. P.; Liu, A.; Pezzicoli, F. S.; Shiba, H.; Volpe, G.; Zamponi, F.; Berthier, L.; and Biroli, G. 2025. Roadmap on Machine Learning Glassy Dynamics. *Nat Rev Phys*, 7(2): 91–104.
- Klushyn, A.; Chen, N.; Kurle, R.; Cseke, B.; and van der Smagt, P. 2019. Learning Hierarchical Priors in VAEs. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.
- Krykunov, M.; and Woo, T. K. 2018. Bond type restricted radial distribution functions for accurate machine learning prediction of atomization energies. ArXiv:1807.10301 [physics].
- Li, H.; Zhang, S.; Tang, L.; Bauchy, M.; and Sun, Y. 2024. Predicting and Interpreting Energy Barriers of Metallic Glasses with Graph Neural Networks. ArXiv:2401.08627.
- Li, K. Y.; Li, M. Z.; and Wang, W. H. 2024. Inverse design machine learning model for metallic glasses with good glass-forming ability and properties. *Journal of Applied Physics*, 135(2): 025102.
- Li, Z.; Mrad, R.; Jiao, R.; Huang, G.; Shan, J.; Chu, S.; and Chen, Y. 2025. Generative Design of Crystal Structures by Point Cloud Representations and Diffusion Model. *iScience*, 28(1): 111659.
- Lin, A.; Huguenin-Dumittan, K. K.; Cho, Y.-C.; Nigam, J.; and Cersonsky, R. K. 2024. Expanding density-correlation machine learning representations for anisotropic coarse-grained particles. *The Journal of Chemical Physics*, 161(7): 074112.
- Long, T.; Fortunato, N. M.; Opahle, I.; Zhang, Y.; Samathrakakis, I.; Shen, C.; Gutfleisch, O.; and Zhang, H. 2021. Constrained Crystals Deep Convolutional Generative Adversarial Network for the Inverse Design of Crystal Structures. *npj Comput Mater*, 7(1): 1–7.
- Luo, X.; Wang, Z.; Gao, P.; Lv, J.; Wang, Y.; Chen, C.; and Ma, Y. 2024. Deep learning generative model for crystal structure prediction. *npj Computational Materials*, 10(1): 1–10. Publisher: Nature Publishing Group.
- McInnes, L.; Healy, J.; and Melville, J. 2020. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv:1802.03426 [stat].
- Mudassir, M. W.; Goverapet Srinivasan, S.; Mynam, M.; and Rai, B. 2022. Systematic Identification of Atom-Centered Symmetry Functions for the Development of Neural Network Potentials. *The Journal of Physical Chemistry A*, 126(44): 8337–8347. Publisher: American Chemical Society.

- Nouira, A.; Sokolovska, N.; and Crivello, J.-C. 2019a. CrystalGAN: Learning to Discover Crystallographic Structures with Generative Adversarial Networks. arXiv:1810.11203.
- Nouira, A.; Sokolovska, N.; and Crivello, J.-C. 2019b. CrystalGAN: Learning to Discover Crystallographic Structures with Generative Adversarial Networks. ArXiv:1810.11203 [cs].
- O’Boyle, N.; and Dalke, A. 2018. DeepSMILES: An Adaptation of SMILES for Use in Machine-Learning of Chemical Structures.
- Pauly, S.; Gorantla, S.; Wang, G.; Kühn, U.; and Eckert, J. 2010. Transformation-mediated ductility in CuZr-based bulk metallic glasses. *Nature Materials*, 9(6): 473–477. Publisher: Nature Publishing Group.
- Rapaport, D. C. 2004. *The Art of Molecular Dynamics Simulation*. Cambridge: Cambridge University Press, 2 edition. ISBN 978-0-521-82568-9.
- Ren, Z.; Tian, S. I. P.; Noh, J.; Oviedo, F.; Xing, G.; Li, J.; Liang, Q.; Zhu, R.; Aberle, A. G.; Sun, S.; Wang, X.; Liu, Y.; Li, Q.; Jayavelu, S.; Hippalgaonkar, K.; Jung, Y.; and Buonassisi, T. 2022. An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. *Matter*, 5(1): 314–335. ArXiv:2005.07609 [physics].
- Sanchez-Gonzalez, A.; Godwin, J.; Pfaff, T.; Ying, R.; Leskovec, J.; and Battaglia, P. W. 2020. Learning to Simulate Complex Physics with Graph Networks. ArXiv:2002.09405 [cs].
- Satorras, V. G.; Hoogeboom, E.; and Welling, M. 2022. E(n) Equivariant Graph Neural Networks. ArXiv:2102.09844 [cs].
- Schleder, G. R.; Padilha, A. C. M.; Acosta, C. M.; Costa, M.; and Fazzio, A. 2019. From DFT to machine learning: recent approaches to materials science—a review. *Journal of Physics: Materials*, 2(3): 032001. Publisher: IOP Publishing.
- Schroers, J.; Hodges, T. M.; Kumar, G.; Raman, H.; Barnes, A. J.; Pham, Q.; and Waniuk, T. A. 2011. Thermoplastic blow molding of metals. *Materials Today*, 14(1): 14–19.
- Skinnider, M. A. 2024. Invalid SMILES are beneficial rather than detrimental to chemical language models. *Nature Machine Intelligence*, 6(4): 437–448. Publisher: Nature Publishing Group.
- Starr, F. W.; Sastry, S.; Douglas, J. F.; and Glotzer, S. C. 2002. What do we learn from the local geometry of glass-forming liquids? *Physical Review Letters*, 89(12): 125501. ArXiv:cond-mat/0206121.
- Sun, H.; Hamel, S.; Hsu, T.; Sadigh, B.; Lordi, V.; and Zhou, F. 2024. Ice phase classification made easy with score-based denoising. ArXiv:2405.06599 [cond-mat] version: 1.
- Trexler, M. M.; and Thadhani, N. N. 2010. Mechanical properties of bulk metallic glasses. *Progress in Materials Science*, 55(8): 759–839.
- Vahdat, A.; and Kautz, J. 2020. NVAE: A Deep Hierarchical Variational Autoencoder. In *Advances in Neural Information Processing Systems*, volume 33, 19667–19679. Curran Associates, Inc.
- Wang, A. Y.-T.; Kauwe, S. K.; Murdock, R. J.; and Sparks, T. D. 2021. Compositionally restricted attention-based network for materials property predictions. *npj Computational Materials*, 7(1): 1–10. Publisher: Nature Publishing Group.
- Wang, Q.; and Zhang, L. 2021. Inverse design of glass structure with deep graph neural networks. *Nature Communications*, 12(1): 5359. Publisher: Nature Publishing Group.
- Watanabe, N.; Hori, Y.; Sugisawa, H.; Ida, T.; Shoji, M.; and Shigeta, Y. 2024. A machine learning potential construction based on radial distribution function sampling. *Journal of Computational Chemistry*, 45(32): 2949–2958. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.27497>.
- Weininger, D. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1): 31–36. Publisher: American Chemical Society.
- Xie, T.; France-Lanord, A.; Wang, Y.; Shao-Horn, Y.; and Grossman, J. C. 2019. Graph dynamical networks for unsupervised learning of atomic scale dynamics in materials. *Nature Communications*, 10(1): 2667. Publisher: Nature Publishing Group.
- Xu, X.; and Hu, J. 2023. A Generative Adversarial Networks (GAN) Based Efficient Sampling Method for Inverse Design of Metallic Glasses. *Journal of Non-Crystalline Solids*, 613: 122378.
- Ying, R.; You, J.; Morris, C.; Ren, X.; Hamilton, W. L.; and Leskovec, J. 2019. Hierarchical Graph Representation Learning with Differentiable Pooling. arXiv:1806.08804.
- Yoshikawa, K.; Yano, K.; Goto, S.; Kim, K.; and Matubayasi, N. 2025. Graph Neural Network-based structural classification of glass-forming liquids and its interpretation via Self-Attention mechanism. ArXiv:2505.00993 [cond-mat].
- Yu, Y.; Wang, M.; Zhang, D.; Wang, B.; Sant, G.; and Bauchy, M. 2015. Stretched Exponential Relaxation of Glasses at Low Temperature. *Physical Review Letters*, 115(16): 165901. Publisher: American Physical Society.
- Yu, Z.; Liu, Q.; Szlufarska, I.; and Wang, B. 2021. Structural signatures for thermodynamic stability in vitreous silica: Insight from machine learning and molecular dynamics simulations. *Physical Review Materials*, 5(1): 015602. Publisher: American Physical Society.
- Zeni, C.; Pinsler, R.; Zügner, D.; Fowler, A.; Horton, M.; Fu, X.; Shysheya, S.; Crabbé, J.; Sun, L.; Smith, J.; Nguyen, B.; Schulz, H.; Lewis, S.; Huang, C.-W.; Lu, Z.; Zhou, Y.; Yang, H.; Hao, H.; Li, J.; Tomioka, R.; and Xie, T. 2024. MatterGen: A Generative Model for Inorganic Materials Design. arXiv:2312.03687.