STRUCTEVAL: A BENCHMARK FOR EVALUATING GENERATION, INFERENCE, AND RECONSTRUCTION IN ATOMIC AND CRYSTALLINE STRUCTURES

Anonymous authors
Paper under double-blind review

ABSTRACT

Crystalline materials power technologies from solar conversion to catalysis, yet current machine learning evaluations artificially divide infinite lattices and finite nanoclusters into separate domains. StructEval unifies these regimes with a symmetry-aware, radius-resolved framework that systematically links primitive unit cells to nanoparticles across ten industrially critical compounds. Each material includes 20+ radii configurations between 0.6–3.0 nm, sampled at 780 quasi-uniform orientations, producing nearly $200,000$ structures spanning $55 - 11,298$ atoms. StructEval defines three rigorous challenges–unit cell $\rightarrow$ nanoparticle generation, nanoparticle $\rightarrow$ unit cell resolution, and lattice reconstruction–supported by a leakage-free split that isolates orientation interpolation from radius extrapolation. This design enables precise measurement of generalization across scale and symmetry. Benchmarking leading generative models reveals severe breakdowns under out-of-distribution conditions, exposing a fundamental gap in current architectures. By providing a reproducible, geometry-grounded testbed fully compatible with PyTorch Geometric, StructEval establishes the foundation for next-generation generative, inference, and reconstruction models in crystalline systems. Data and implementations are released at `https://anonymous.4open.science/r/StructEval-ANONYMOUS`.

1 INTRODUCTION

Nanostructured materials enable applications from photovoltaic energy conversion to electronic sensing Simonov & Goodwin (2020). Their behavior depends not only on the periodic arrangement defined by a primitive unit cell but also on the finite morphologies and surface terminations that arise in nanoscale particles Cao et al. (2022). These two regimes—primitive cells and nanoclusters—have traditionally been studied in isolation. Bottom-up crystal growth simulations typically begin with an ideal lattice and sequentially add atoms, whereas nanoparticle modeling workflows construct finite clusters and optimize them using empirical and first-principles methods Levi & Kotrla (1997). Such approaches are computationally intensive and difficult to scale across compositions, particle sizes, and orientations Surek (2005). Ab initio electronic-structure methods, particularly density functional theory (DFT) Orio et al. (2009) and its tight-binding approximation (DFTB) Elstner & Seifert (2014), are widely used to refine geometries, compute formation energies, and predict surface reconstructions. While DFT provides high accuracy, its steep computational cost inhibits large-scale exploration Cohen et al. (2008). DFTB mitigates this cost by approximating the DFT Hamiltonian with a compact basis, enabling efficient evaluation of nanoparticle energetics across many sizes and orientations Liu et al. (2019); Qi et al. (2013).

Machine learning (ML) offers an alternative path by learning directly from data rather than relying exclusively on physics-based computationally heavy simulations Carleo et al. (2019); Karniadakis et al. (2021); Alpaydin (2021). Graph neural networks (GNNs) such as SchNet Schütt et al. (2018), DimeNet++ Gasteiger et al. (2020), TorchMD-Net Thölke & De Fabritiis (2022), and CGNNs Cheng et al. (2021) have demonstrated strong performance in predicting properties of molecules and crystals, using datasets such as QM9 Ruddigkeit et al. (2012); Ramakrishnan et al. (2014) and MD17 Chmiela et al. (2017). Developments in equivariant architectures—such as E(n)-GNNs Satorras et al. (2021), SE(3)-Transformers Fuchs et al. (2020), PaiNN Schütt et al. (2021), SphereNet Coors et al. (2018), Equiformer Liao & Smidt (2022), FAENet Duval et al. (2023), and GotenNet Aykent & Xia (2025)—improve data efficiency by enforcing geometric symmetries. Multimodal architectures like Pure2DopeNet Polat et al. (2024), MolPROP Rollins et al. (2024), and CrysMMNet Das et al.

(2023) extend this further by integrating multiple input types. Generative models such as CDVAE Xie et al. (2021), GraphDF Luo et al. (2021), UniMat Yang et al. (2023), and CrystalFlow Luo et al. (2024) show promise in generating crystal structures via score-based diffusion and flow-based modeling. However, evaluation of these models is typically limited to small molecules or bulk crystals under in-distribution settings. Despite this progress, there is no benchmark that systematically couples primitive-cell inputs to radius- and orientation-controlled nanoparticle structures. None offer symmetry-driven nanoparticle manifolds tied directly to bulk lattices, nor do they support robust evaluation under generalization to new sizes or spatial orientations. As a result, model development has largely proceeded without clarity on how generative or inference models behave under realistic distribution shifts in scale and symmetry.

To address this gap, we introduce StructEval, a symmetry-aware, radius-resolved dataset that links primitive unit cells to systematically generated nanoparticle structures. For each of ten materials pivotal to engineering applications (Ag, Au, $CH_3NH_3PbI_3$, Fe_2O_3, MoS_2, PbS, SnO_2, $SrTiO_3$, TiO_2, ZnO), dataset provides: (i) a primitive unit cell, (ii) 25 nanoparticle radii ranging from 0.6 to 3.0 nm, and (iii) 780 deterministic quasi-uniform orientations per radius. This results in about 200,000 atomic configurations, spanning 55 to 11,298 atoms per structure. All structures are generated with reproducible, leakage-free protocols that support controlled evaluation of orientation and scale generalization. StructEval defines three benchmark tasks that span both predictive and generative modeling: (1) unit cell $\rightarrow$ nanoparticle generation, (2) nanoparticle $\rightarrow$ lattice inference, and (3) structural reconstruction from partial observations. A leakage-free split strategy separates interpolation over orientations (ID) from extrapolation across radii (OOD), enabling rigorous benchmarking of symmetry- and scale-aware models. Evaluations of state-of-the-art (SOTA) generative architectures — including DiffCSP Jiao et al. (2023), FlowMM Miller et al. (2024), FlowLLM Sriram et al. (2024), MatterGen-MP Zeni et al. (2023), and ADiT Joshi et al. (2025) — on all tasks reveal consistent performance degradation under OOD splits. StructEval fills a critical gap in the landscape of atomistic ML benchmarks by unifying primitive symmetry and nanoparticle structure within a single, reproducible framework. By supporting multiple tasks and emphasizing distribution shifts, StructEval enables systematic evaluation of generative and inference models in crystalline systems.

The remainder of the paper is organized as follows. Section 2 presents a review of prior work on crystalline building blocks and nanostructured morphologies, including first-principles and semi-empirical simulation methods, as well as existing benchmarks for crystals and nanomaterials. Section 3 details the construction and characteristics of the StructEval dataset. Section 4 discusses the performance of the SOTA generative models on StructEval across three distinct tasks. Section 5 outlines the current limitations of StructEval dataset. Section 6 summarizes the key findings and proposes directions for future research.

2 Related Work

2.1 Crystalline Building Blocks and Nanostructured Morphologies

The primitive unit cell defines the translational symmetry, atomic motifs, and electronic band topology that govern bulk-scale material behavior Kittel & McEuen (2018). When extended crystals are truncated into finite clusters, their equilibrium morphologies are determined by orientation-dependent surface free energies, as described by the Gibbs–Wulff theorem Li et al. (2016). This leads to well-defined faceted shapes, such as truncated octahedra for face-centered cubic metals and coherent morphologies in perovskite structures Barmparis et al. (2015); Mattoni et al. (2016); Najman et al. (2021). Experimental and first-principles studies have validated the predictive power of the Wulff construction across metals, oxides, and hybrid perovskites Ringe et al. (2013). Below radii of approximately 3 nm, quantum confinement effects shift electronic states into discrete levels, significantly altering optical gaps, plasmonic responses, and catalytic properties. These size-dependent phenomena have been extensively characterized in metallic quantum dots and chalcogenide nanocrystals Bera et al. (2010). Such characteristics enable a range of applications, including localized surface plasmon biosensors based on Ag and Au nanocubes, high-surface-area TiO_2 photocatalysts, perovskite-based light emitters, and ZnO piezotronic sensors Di Fonzo et al. (2008); Chen et al. (2024). Despite the technological importance of nanocrystalline morphology, a systematic dataset mapping primitive-cell symmetries to radius-resolved nanoclusters is currently lacking. This gap hinders the development and evaluation of ML models capable of bridging crystalline behavior across length scales Yang et al. (2022).

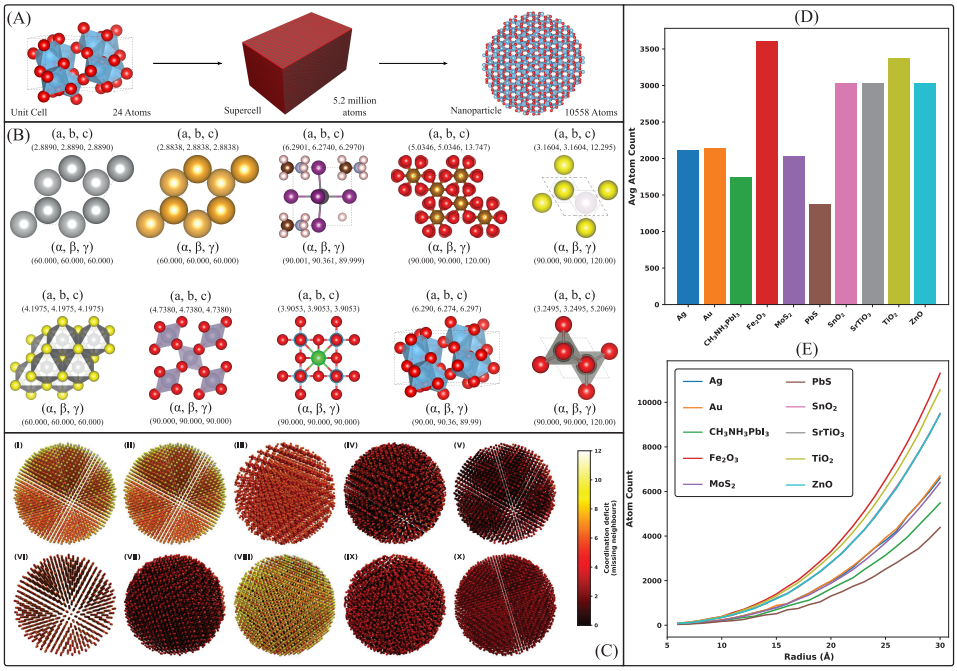

Figure 1: Detailed overview of StructEval. (A) Workflow for generating radius-resolved nanoparticles. (B) Unit cells of the materials—Ag, Au, $CH_3NH_3PbI_3$, Fe_2O_3, MoS_2, PbS, SnO_2, $SrTiO_3$, TiO_2, and ZnO—arranged left to right. The lattice constants a, b and c denote the cell edge lengths, while the angles α, β and γ specify the inter-edge angles between b–c, a–c and a–b, respectively. (C) Coordination-deficit heatmaps for the largest nanoparticle of each material (panels I–X follow the same material order). (D) Mean atom count per material, illustrating typical nanoparticle sizes. (E) Atom count versus nanoparticle radius, revealing how total atom number scales with size.

2.2 First-Principles and Semi-Empirical Simulation Methods

Kohn–Sham DFT delivers accurate predictions of lattice constants, cohesive energies, and surface reconstructions, but its $\mathcal{O}(N^3)$ diagonalization cost limits feasible supercell sizes to approximately 10^3 atoms on conventional hardware Bickelhaupt & Baerends (2000); Yu et al. (2016). Linear-scaling formulations such as ONETEP reduce this cost by exploiting density-matrix locality, enabling simulations with tens of thousands of atoms while retaining plane-wave accuracy, even for large nanorods and complex oxides Baer & Head-Gordon (1997); Skylaris et al. (2005). To explore broader compositional and orientational spaces, semi-empirical approaches are essential. DFTB approximates the Kohn–Sham Hamiltonian using a minimal basis, providing two to three orders of magnitude speed-up, which allows high-throughput screening of nanoparticle stability, defect segregation, and vibrational spectra Zheng et al. (2005); Spiegelman et al. (2020); Bačić et al. (2020); Kim et al. (2019). Classical interatomic potentials—such as the embedded atom method and machine-learned force fields—further extend accessible system sizes into the tens of nanometers for metals and alloys, though at the cost of reduced chemical accuracy and transferability Daw & Baskes (1984); Mahata et al. (2022); Behler & Parrinello (2007). Each modeling tier thus presents a trade-off between accuracy and computational throughput, motivating the development of ML surrogates that aim to combine quantum-level fidelity with the scalability of empirical methods Kurban et al. (2024).

2.3 Benchmarks for Crystals and Nanomaterials

Benchmark datasets have driven significant advances in applying ML to quantum chemistry and materials science by providing structured data for diverse predictive tasks. In addition to the ones mentioned in Section 1, QM7 Blum & Reymond (2009); Rupp et al. (2012) targets atomization energies of small organic molecules, while MD22 Chmiela et al. (2023) offers molecular dynamics trajectories labeled with atomic forces for force-field model training. PubChemQC Kim et al. (2025) extends this to millions of ground-state molecular geometries with corresponding electronic properties. Datasets such as NablaDFT Khrabrov et al. (2022) and QH9 Yu et al. (2024) focus on Hamiltonian matrix estimation, with NablaDFT contributing extensive data across molecular

conformers. Perov-5 Castelli et al. (2012a;b) includes 18,928 cubic ABX_3 perovskites drawn from an open water-splitting DFT database, while Carbon-24 Pickard (2020) provides over 10,000 distinct low-energy carbon frameworks identified via AIRSS at 10 GPa and relaxed using DFT. In the domain of inorganic solids, the Materials Project underpins MatBench Jain et al. (2020), which hosts fourteen property-prediction benchmarks—including band gaps, bulk moduli, and elastic tensors. Surface energetics are addressed by the Open Catalyst datasets: OC20 Chanussot et al. (2021) connects relaxed metal-surface structures with atomic forces, and OC22 Tran et al. (2023) expands coverage to oxide slabs. More recent initiatives such as LAMBench Peng et al. (2025) aggregate varied geometries for large-scale atomic model pretraining, and CrysMTM Polat et al. (2025) introduces temperature-dependent nanomaterial configurations with multimodal annotations. While these benchmarks offer broad coverage, most focus exclusively on either infinite bulk crystals or small, gas-phase molecules. Radius-graded nanoclusters—structures intermediate in scale and symmetry—typically appear only as side products in catalytic or surface simulations. Temperature dependence, phase transitions, and symmetry-aware design are also often underrepresented. To date, no public benchmark systematically links primitive-cell representations with radius- and orientation-resolved nanoclusters across diverse chemistries and size regimes. Furthermore, none explicitly supports bidirectional modeling tasks such as nanoparticle generation and lattice inference from finite clusters. This gap motivates the development of StructEval.

3 Dataset Construction and Task Definition

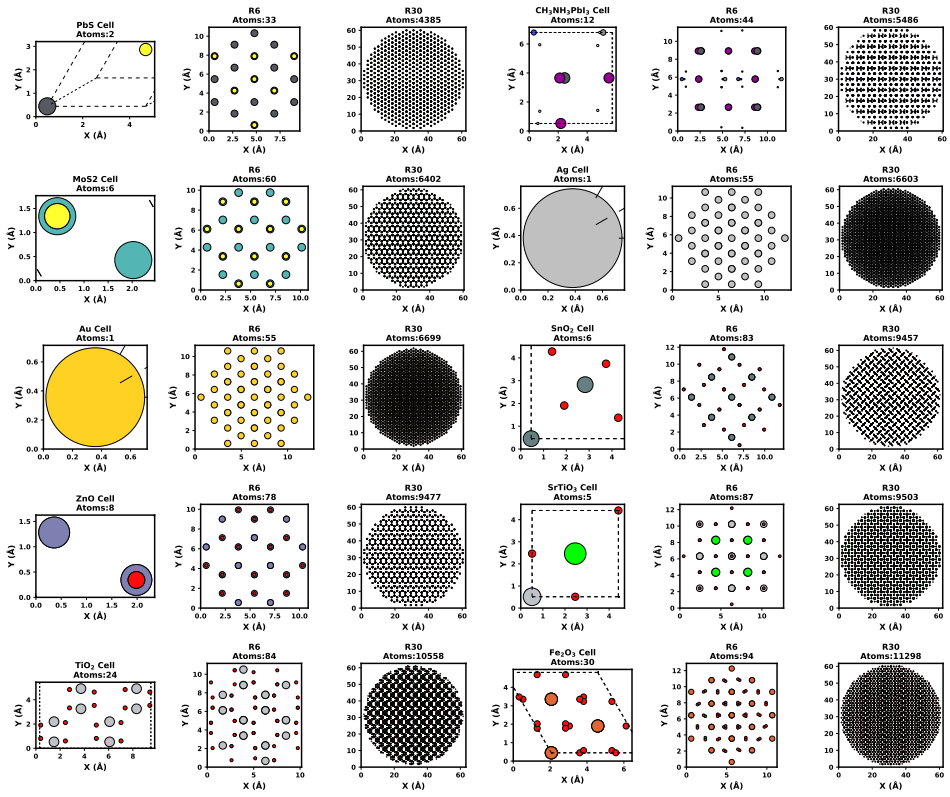

Figure 2: From primitive cell to radius-controlled nanoclusters. For each material in dataset, the panels show—left to right—the primitive unit cell followed by its canonical R = 6 Å and R = 30 Å nanoparticles. Materials are arranged from top to bottom in ascending order of the atom count in their R30 cluster, illustrating how coordination environments and bulk-like cores emerge with increasing radius. All views share a common Ångström scale. Atom colours follow the conventional CPK palette: Ag (light grey), Au (gold), C (dark grey), H (white), N (blue), O (red), S (yellow), Fe (orange-brown), Mo (teal), Sn (silver), Sr (green), Ti (slate grey), Zn (forest green), Pb (dark grey-blue), and I (purple).

3.1 CRYSTAL STRUCTURES AND NANOPARTICLE GENERATION

The STRUCTEVAL suite spans elemental solids, perovskites, transition-metal dichalcogenides, and binary oxides. The selected compounds include Silver (Ag) King (2002b), Gold (Au) King (2002a), methylammonium lead iodide ($CH_3NH_3PbI_3$) Walsh et al. (2019), hematite (Fe_2O_3) Finger & Hazen (1980), molybdenum disulfide (MoS_2) Wyckoff (1963a); Grau-Crespo & Lopez-Cordero (2002), galena (PbS) Wyckoff (1963b), cassiterite (SnO_2) Baur et al. (1971), strontium titanate ($SrTiO_3$) Mitchell & Carpenter (2000), anatase titanium dioxide (TiO_2) Horn et al. (1972), and zincite zinc oxide (ZnO) Wyckoff (1963c). Primitive unit cells were extracted from CIFs and serve as the basis for nanoparticle generation and symmetry-aware modeling tasks. Lattice parameters, space groups, and atomic positions are provided in Appendix B, while Figure 2 summarizes dataset structure and key statistics.

Supercell construction. Let $N = 60$. We define the periodic lattice of atomic sites as

$$\Lambda \;=\; \Big\{ n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 + \mathbf{r}_j \;\Big|\; n_1, n_2, n_3 \in \{0, \dots, N-1\},\; j \in \{1, \dots, M\} \Big\}, \quad (1)$$

where $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \in \mathbb{R}^3$ are the primitive lattice vectors of the unit cell, and $\{\mathbf{r}_j\}_{j=1}^{M} \subset \mathbb{R}^3$ are the basis positions of the M atoms inside each unit cell. The integers $n_1, n_2, n_3 \in \{0, 1, \dots, N-1\}$ serve as lattice indices, specifying the unit cell position along the three lattice directions, and may equivalently be viewed as elements of $\mathbb{Z}_N$. Replicating the unit cell N times along each direction yields an $N \times N \times N$ supercell, and the set Λ in equation 1 collects all atomic sites contained within it.

Reference atom. Choose a central reference position $\mathbf{r}_0 \in \Lambda$ (e.g., the image of $\mathbf{r}_1$ in the central unit cell defined by equation 1).

Nanoparticle definition. For a given cutoff radius R, define the nanoparticle as

$$P(R) \;=\; \Big\{ \mathbf{x} \in \Lambda \;\Big|\; \|\mathbf{x} - \mathbf{r}_0\|_2 \le R \Big\}, \quad (2)$$

i.e., all lattice sites within a sphere of radius R around $\mathbf{r}_0$.

Radius sampling. Select $K = 25$ radii on a uniform grid:

$$R_k \;=\; R_{\min} + (k-1)\,\Delta R, \qquad k = 1, \dots, K, \quad (3)$$

with end-points and step size

$$R_{\min} = 0.6\,\text{nm}, \qquad R_{\max} = 3.0\,\text{nm}, \qquad \Delta R \;=\; \frac{R_{\max} - R_{\min}}{K-1} \;=\; 0.1\,\text{nm}. \quad (4)$$

Relationships captured by the equations 2–4 define the complete set $\{P(R_k)\}_{k=1}^{25}$, spanning quantum-confined clusters ($R \le 1.0\,\text{nm}$) to near-bulk fragments ($R \ge 2.0\,\text{nm}$).

3.2 QUASI-UNIFORM ROTATIONS

Deterministic, quasi-uniform rotations ensure that each nanoparticle orientation is represented without randomness, enabling reproducible data augmentation and unbiased evaluation of rotation-equivariant models. By constructing a finite, low-discrepancy *set* on $\mathrm{SO}(3)$ that includes the identity and approximates uniform coverage, one can systematically generate rotated copies of any atomic configuration Yershova et al. (2010). Let $\mathcal{P} = \{\mathbf{r}_1, \dots, \mathbf{r}_n\} \subset \mathbb{R}^3$ denote a finite point cloud (with n points). Our goal is to build a finite set $\mathcal{G} \subset \mathrm{SO}(3)$ containing the identity that provides near-uniform, reproducible sampling of rotation space. The construction proceeds in three steps:

Axis–angle parametrization. Every rotation $R \in \mathrm{SO}(3)$ is represented as

$$R \;=\; R(\hat{\mathbf{u}}, \phi), \quad (5)$$

where $\hat{\mathbf{u}} \in \mathbb{S}^2$ is a unit axis and $\phi \in [0, \pi]$ is the rotation angle, with the equivalence $(\hat{\mathbf{u}}, \phi) \sim (-\hat{\mathbf{u}}, 2\pi - \phi)$ ensuring a unique representation. Using Rodrigues' formula Dai (2015), this rotation is expressed as

$$R(\hat{\mathbf{u}}, \phi) = \; I_3 \cos\phi + (1 - \cos\phi)\,\hat{\mathbf{u}}\,\hat{\mathbf{u}}^\top + \sin\phi\,[\hat{\mathbf{u}}]_\times, \quad (6)$$

where I_3 is the 3×3 identity matrix, $\hat{\mathbf{u}} = (u_1, u_2, u_3)^\top$, and

$$[\hat{\mathbf{u}}]_\times \;=\; \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \quad (7)$$

is the skew-symmetric matrix implementing the linear map $\mathbf{v} \mapsto \hat{\mathbf{u}} \times \mathbf{v}$, which guarantees that $R(\hat{\mathbf{u}}, \phi)$ in equation 6 is orthogonal and has determinant 1.

Quasi-Uniform Grid on $\mathrm{SO}(3)$. We define a deterministic sampling scheme via the Hopf fibration Lyons (2003):

$$(\hat{\mathbf{u}}, \psi) \;\longmapsto\; R\big(\hat{\mathbf{u}},\, 2\psi\big), \qquad \hat{\mathbf{u}} \in \mathbb{S}^2,\ \psi \in [0, \pi).$$

To discretize this space, we proceed as follows.

1. **Fibonacci lattice on $\mathbb{S}^2$.** For $k = 0, \ldots, N_{\text{axis}} - 1$ set

$$z_k = 1 - \frac{2k+1}{N_{\text{axis}}}, \quad r_k = \sqrt{1 - z_k^2}, \quad \varphi_g = \frac{1+\sqrt{5}}{2}, \quad \varphi_k = 2\pi k\, \varphi_g^{-1},$$

   and define the axis $\hat{\mathbf{u}}_k = (r_k \cos\varphi_k,\ r_k \sin\varphi_k,\ z_k)$.

2. **Irrational angle increments.** For $m = 0, \ldots, M_{\text{ang}} - 1$, let

$$\psi_m \;=\; 2\pi m\, \varphi_g^{-1} \mod 2\pi.$$

3. **Rotation-set construction.** Define $R_{k,m} = R\big(\hat{\mathbf{u}}_k,\, 2\psi_m\big)$ and let $\mathcal{G}$ be the set consisting of the identity and all $R_{k,m}$. Because $\psi_0 = 0$ yields $R = I$, we count identity once, hence

$$|\mathcal{G}| \;=\; 1 + N_{\text{axis}}\,(M_{\text{ang}} - 1).$$

Quasi-Uniform Coverage (sketch). As $N_{\text{axis}}, M_{\text{ang}} \to \infty$, the induced empirical measure of the above low-discrepancy set provides increasingly uniform coverage of $\mathrm{SO}(3)$ with respect to the Haar measure (in the sense of vanishing discrepancy), leveraging the equidistribution of the Fibonacci lattice on $\mathbb{S}^2$ Stanley (1975) and the incommensurate twist increments John (1998).

Rigid-body action on the point cloud. Let $\mathcal{P} = \{\mathbf{r}_1, \ldots, \mathbf{r}_n\} \subset \mathbb{R}^3$ be a finite point cloud and $\mathbf{c} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{r}_i$ be its center of mass. For each $R \in \mathcal{G}$, define the rigidly rotated configuration

$$\mathcal{P}_R \;=\; \big\{\, R(\mathbf{r}_i - \mathbf{c}) + \mathbf{c} \;\big|\; i = 1, \ldots, n \big\}, \tag{8}$$

which preserves all inter-point distances while re-centering about $\mathbf{c}$. To ensure numerical robustness, a tolerance $\varepsilon > 0$ is used when applying equation 8: near-zero angles are snapped, degenerate axes are regularized, and infinitesimal perturbations break floating-point degeneracies. This deterministic augmentation yields a near-uniform sampling of orientations from $\mathcal{G}$ without introducing stochastic noise.

Configuration diversity. Although individual nanoparticles may exhibit high symmetry, the Cartesian product of (i) ten distinct chemistries/crystal families, (ii) 25 radii spanning quantum-confined to near-bulk regimes, and (iii) $|\mathcal{G}| = 780$ deterministic orientations produces $\approx 200{,}000$ unique atomic configurations, capturing rich variation in surface terminations, coordination statistics, facet exposure, and global morphology—especially for non-cubic/anisotropic lattices.

3.3 Data Splits

For each material m, define the set of labeled nanoparticle instances as

$$\mathcal{G}_m \;=\; \big\{\, (R, b) \mid R \in \{\mathrm{R6}, \ldots, \mathrm{R30}\},\ b \in \{0, \ldots, 780\} \big\}, \tag{9}$$

which enumerates every radius–orientation pair for material m. This space is partitioned into

$$\mathcal{G}_m^{\text{in}} \;=\; \big\{(R, b) \in \mathcal{G}_m \mid R \le \mathrm{R24}\big\}, \qquad \mathcal{G}_m^{\text{out}} \;=\; \mathcal{G}_m \setminus \mathcal{G}_m^{\text{in}}. \tag{10}$$

A random split (without replacement) is applied to $\mathcal{G}_m^{\text{in}}$, yielding disjoint subsets $\mathcal{T}_m$ (training, 70 %), $\mathcal{V}_m$ (validation, 10 %), and $\mathcal{I}_m$ (ID test, 20 %). The resulting evaluation partitions are therefore $\mathsf{Train}_m = \mathcal{T}_m$, $\mathsf{Val}_m = \mathcal{V}_m$, $\mathsf{IDTest}_m = \mathcal{I}_m$, and $\mathsf{OODTest}_m = \mathcal{G}_m^{\text{out}}$, all defined consistently via equations 9 and 10.

Grouping every orientation (R, b) together ensures each split is rotation-invariant, preventing orientation leakage. To further test rotational equivariance, a fixed phase shift is applied to the index b in $\mathsf{OODTest}_m$: every OOD b value is offset by a constant Δb, guaranteeing that no rotation present in the training, validation, or ID-test sets reappears in the OOD set. This choice—implemented as a 15° rotation about the non-degenerate $\langle 111 \rangle$ axis—keeps the shifted and unshifted grids provably disjoint under machine precision, thereby isolating orientation interpolation from scale extrapolation.

Physical grounding and scope. Nanoparticles in this benchmark are carved directly from experimentally characterized bulk crystals using a deterministic spherical truncation pipeline, thereby

preserving the local crystallographic environment while intentionally breaking global periodicity. We do not label structures with thermodynamic or electronic properties (e.g., formation energies, band gaps), nor do we perform ab-initio relaxations at scale; this is a deliberate design choice to keep the benchmark geometry- and symmetry-focused, clean, and fully reproducible at $\approx 200,000$ configurations. Given the cubic scaling of DFT with system size and that our largest clusters reach $\approx 10^3$–10^4 atoms, exhaustive first-principles labeling is computationally infeasible; instead, we emphasize a deterministic, symmetry-aware construct that isolates structural reasoning and supports transparent ID/OOD splits.

3.4 TASKS

The dataset encompasses three core tasks, each paired with multiple evaluation metrics that capture both geometric and crystallographic aspects. These include: root-mean-square deviation (RMSD), $\mathrm{RMSD}(P, P^*)$; Hausdorff distance, $d_{\mathrm{Haus}}(P, P^*)$ Rucklidge (1997); convex hull volume ratio error, $\Delta V_{\mathrm{hull}}(P, P^*) = |\mathrm{V}_{\mathrm{hull}}(P)/\mathrm{V}_{\mathrm{hull}}(P^*) - 1|$; and radial distribution function divergence, $E_{\mathrm{RDF}}(P, P^*) = \int |g_P(r) - g_{P^*}(r)|\, \mathrm{d}r$. Validity rate is measured via $\mathrm{V}_{\mathrm{R}}(P)$. In addition, Task 2 employs the root mean square error of lattice parameters, $\mathrm{RMSE}(\ell, \ell^*)$, and space-group accuracy $\mathbb{1}[g = g^*]$ along with joint accuracy $\mathbb{1}[(\ell, g) = (\ell^*, g^*)]$. Task 3 further incorporates surface atom recall $\mathrm{R}_{\mathrm{surf}}(P, P^*)$ and coordination histogram KL divergence $D_{\mathrm{KL}}\big(h(P) \,\|\, h(P^*)\big)$. All three tasks use the splits Train_m, Val_m, IDTest_m, and $\mathsf{OODTest}_m$ to separately quantify orientation interpolation and scale extrapolation for each material m. An overview of the tasks is illustrated in Figure 3 (A), (B), and (C). Extended definitions of these metrics are provided in Appendix C.

Task 1: Unit Cell → Nanoparticle Generation. From a primitive unit cell u_m and a target radius R, the model must generate realistic nanoparticles that faithfully reproduce both the infinite lattice periodicity and the finite-size surface morphology. Mastering this capability is essential for producing computationally efficient nanostructure models that accurately capture surface-driven phenomena, thereby enhancing predictive accuracy in material design:

$$f_1 : (u_m, R) \;\longmapsto\; P \subset \mathbb{R}^3. \tag{11}$$

The mapping in equation 11 is assessed with the metrics $\mathrm{RMSD}(P, P^*)$, $d_{\mathrm{Haus}}(P, P^*)$, $\Delta V_{\mathrm{hull}}(P, P^*)$, $E_{\mathrm{RDF}}(P, P^*)$, and $\mathrm{V}_{\mathrm{R}}(P)$.

Task 2: Nanoparticle → Lattice Inference. Local geometric information extracted from finite nanoparticles is used to infer their lattice constants and symmetry group, exploring whether such short-range cues suffice to reconstruct extended periodicity and crystallographic invariants. Achieving this capability is vital for bridging localized observations with global structural order, thereby enabling accurate property predictions and reliable materials-design workflows:

$$f_2 : P \;\longmapsto\; (\ell, g), \quad \ell = (a, b, c, \alpha, \beta, \gamma), \quad g \in \mathcal{S}, \tag{12}$$

so that equation 12 maps each nanoparticle P to its lattice parameters ℓ and a space-group label g drawn from the set $\mathcal{S}$ of crystallographic groups. Evaluation metrics are then $\mathrm{RMSE}(\ell, \ell^*)$, $\mathbb{1}[g = g^*]$, and the joint accuracy $\mathbb{1}[(\ell, g) = (\ell^*, g^*)]$, all assessed with respect to the ground-truth pair (ℓ^*, g^*).

Task 3: Lattice Reconstruction. Partial occlusion of nanoparticles—whether applied at random or biased toward surface atoms—tests the model's capability to reconstruct the complete atomic arrangement from incomplete inputs. Such evaluation reveals whether the model can infer missing structural information and preserve chemical fidelity, a critical requirement for reliable nanoparticle modeling and downstream materials-design applications:

$$f_3 : P_{\mathrm{mask}} \mapsto P, \quad P_{\mathrm{mask}} = M \cdot P, \tag{13}$$

where M in Eq. (13) is a masking operator that removes a fraction p of atoms. Evaluation metrics include $\mathrm{RMSD}(P, P^*)$, $\mathrm{R}_{\mathrm{surf}}(P, P^*)$, $D_{\mathrm{KL}}\big(h(P) \,\|\, h(P^*)\big)$, and $\mathrm{V}_{\mathrm{R}}(P, P^*)$.

4 EXPERIMENTS

Multiple generative models— such as DiffCSP, FlowMM, FlowLLM, MatterGen-MP, and ADiT—are evaluated on all three tasks using the splits defined in Section 3.4. StructEval has been adapted to be compatible with the PyTorch Geometric Fey & Lenssen (2019) library and the model implementations are developed using PyTorch Paszke (2019) and model's respective official libraries. The details of the implementation and the model settings are shared in Appendix D while comprehensive experimental results for each task are presented in Appendix E.

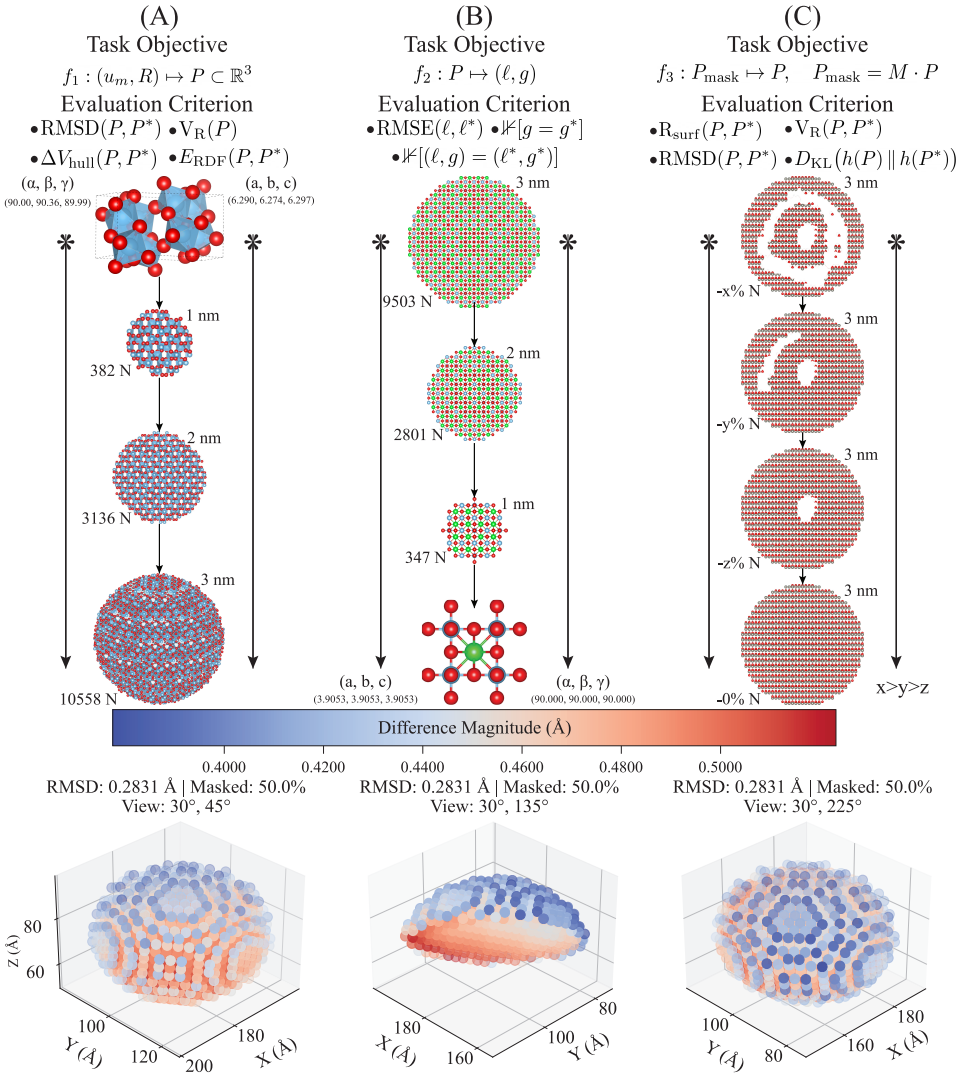

Figure 3: Overview of benchmark tasks and lattice reconstruction experiment result. (A) shows generation of nanoparticles starting from the unit cell. (B) includes the generating unit cell configurations from the nanoparticles. Lastly, (C) illustrates lattice reconstruction from partially masked configurations. N is the atom count in the system. First material is TiO_2, second is $SrTiO_3$ while the last one is ZnO. Scatter plots in the bottom are different angle reconstruction results from CDVAE model for a single material on OOD set with masked 50% of the original atoms.

4.1 Task 1: Unit Cell to Nanoparticle Generation

Quantitative analysis reveals marked performance degradation on OOD nanoparticles. On the ID test, reconstruction loss and KL divergence are 0.0088 and 0.7888 (total 0.7976); on OOD they rise to 0.0788 and 80.1785 (total 80.2573), an $\approx 100\times$ increase. RMSD increases from 0.0671 Å to 0.2295 Å (+242 %), and Hausdorff distance from 0.1597 Å to 0.4783 Å (+200 %). Convex hull deviation jumps $31.15 \rightarrow 178.60$ (+472 %), and radial distribution error $1.1389 \rightarrow 1.3910$ (+22 %). Although the validity ratio climbs $81.98\% \rightarrow 93.90\%$, geometric fidelity suffers. These trends highlight the need for stronger inductive biases (e.g., symmetry constraints or hierarchical modeling) and targeted data-augmentation or domain-adaptation to bridge the generalization gap.

To clarify the source of OOD difficulty, our evaluation was intentionally configured to measure native generalization rather than adaptation: we use leakage-free OOD splits that simultaneously exclude entire radius ranges and apply disjoint orientation grids, and we evaluate models without task-specific fine-tuning or augmentations. This design isolates how current architectures extrapolate across scale (radius) and symmetry (orientation) when given only primitive-cell input and a target

radius. The consistent OOD degradation across diverse model families suggests a structural limitation—insufficient scale awareness, lack of multi-scale feature hierarchies linking local neighborhoods to global morphology, and weak long-range symmetry constraints—rather than a quirk of our data. These results are diagnostic, indicating that closing the gap will require explicit radius conditioning, hierarchical representations, and geometric inductive biases designed to maintain global coherence.

4.2 Task 2: Nanoparticle to Lattice Inference

On the ID test, the model records an RMSE of 55.240 Å, space-group accuracy of 29.6%, and joint accuracy of 0.0%; on OOD these are 55.246 Å ($+0.01\%$), 29.1% (-1.7%), and 0.0%, respectively. Because each material's space group is constant across all radii, the classifier effectively learns a single per-material label—hence the minimal space-group accuracy change—and the near-identical RMSE shows lattice-parameter recovery generalizes even to larger clusters. The persistent 0.0% joint accuracy, however, highlights the challenge of meeting both precise geometric reconstruction and exact symmetry classification simultaneously.

Note that each nanoparticle retains local symmetry but loses global periodicity due to spherical truncation, making joint recovery of lattice parameters + space group fundamentally ambiguous from finite fragments. The observed 0.0% joint accuracy is therefore not merely a training artifact; it reflects the intrinsic challenge of inferring crystallographic invariants from non-periodic observations—precisely the frontier Task 2 is designed to probe.

4.3 Task 3: Lattice Reconstruction

On the ID test, the model achieves an RMSD of 0.481 Å; on OOD data, RMSD rises to 0.722 Å—a 50 % increase. Surface-overlap and local-coordination metrics show only marginal variation between ID and OOD, masking the sharp jump in atomic-placement errors. This gap highlights the intrinsic challenge of completing missing structures across size regimes and reinforces StructEval's value as a stringent benchmark for driving advances in partial-structure completion. An example reconstruction error map is shared in the bottom part of Figure 3 using scatter plots.

5 Limitations

While StructEval bridges bulk symmetry and finite nanoclusters, it has limitations: it covers only ten materials currently; our clusters are crystallographically grounded fragments—deterministic truncations of experimentally validated bulk structures—but they are not thermodynamically validated or passivated, and they omit temperature, disorder, and environmental effects. This is intentional: by decoupling geometric/symmetry reasoning from energetic confounders, the benchmark provides a clean, deterministic testbed for generation, inference, and reconstruction under strict ID/OOD controls. We acknowledge that experimental nanoparticles can undergo surface reconstructions and environment-dependent stabilization.

6 Conclusion and Future Work

Existing benchmarks emphasize either periodic crystals or small molecules, leaving the transitional bulk-to-nanoscale regime largely unaddressed. StructEval fills this gap with a radius-resolved, symmetry-aware dataset that links primitive unit cells to nanoparticles sampled across 25 radii and 780 quasi-uniform orientations. This unified framework supports three bidirectional tasks—generation, inference, and reconstruction—under strict leakage-free splits, providing a rigorous platform for probing both interpolation and extrapolation.

Across tasks, the state-of-the-art models perform reasonably in-distribution but collapse out-of-distribution, where reconstruction errors, symmetry misclassifications, and completion failures sharply increase. Since DFT scales cubically with system size, simulating large nanoparticles directly is impractical; surrogate models that fail to generalize cannot replace it for predicting lattice parameters or morphologies. These results highlight the limits of local geometric cues and motivate new approaches with explicit symmetry constraints, multi-scale hierarchies, and augmentation strategies. StructEval is designed as a modular platform, with planned extensions to molecular crystals, perovskites, polymers, and multi-phase systems. By unifying bulk and nanoscale regimes, it lays a foundation for developing models that capture the full structural complexity of crystalline nanomaterials.

REFERENCES

Ethem Alpaydin. *Machine Learning*. MIT press, 2021.

Sarp Aykent and Tian Xia. GotenNet: Rethinking Efficient 3D Equivariant Graph Neural Networks. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=5wxCQDtbMo.

Vladimir Bačić, Thomas Heine, and Agnieszka Kuc. Analytical approach to phonon calculations in the scc-dftb framework. *The Journal of Chemical Physics*, 153(14), 2020.

Roi Baer and Martin Head-Gordon. Sparsity of the density matrix in kohn-sham density functional theory and an assessment of linear system-size scaling methods. *Physical Review Letters*, 79(20): 3962, 1997.

Georgios D Barmparis, Zbigniew Lodziana, Nuria Lopez, and Ioannis N Remediakis. Nanoparticle shapes by using wulff constructions and first-principles calculations. *Beilstein Journal of Nanotechnology*, 6(1):361–368, 2015.

W. H. Baur, R. A. Sass, et al. The rutile structure of sno_2. *Acta Crystallographica Section B*, 27:2133, 1971.

Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters*, 98(14):146401, 2007.

Debasis Bera, Lei Qian, Teng-Kuan Tseng, and Paul H Holloway. Quantum dots and their multimodal applications: a review. *Materials*, 3(4):2260–2345, 2010.

F Matthias Bickelhaupt and Evert Jan Baerends. Kohn-sham density functional theory: predicting and understanding chemistry. *Reviews in Computational Chemistry*, pp. 1–86, 2000.

L. C. Blum and J.-L. Reymond. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.*, 131:8732, 2009.

Shuang Cao, Ning Sui, Peng Zhang, Tingting Zhou, Jinchun Tu, and Tong Zhang. Tio2 nanostructures with different crystal phases for sensitive acetone gas sensors. *Journal of Colloid and Interface Science*, 607:357–366, 2022.

Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002, 2019.

Ivano E Castelli, David D Landis, Kristian S Thygesen, Søren Dahl, Ib Chorkendorff, Thomas F Jaramillo, and Karsten W Jacobsen. New cubic perovskites for one-and two-photon water splitting using the computational materials repository. *Energy & Environmental Science*, 5(10):9034–9043, 2012a.

Ivano E Castelli, Thomas Olsen, Soumendu Datta, David D Landis, Søren Dahl, Kristian S Thygesen, and Karsten W Jacobsen. Computational screening of perovskite metal oxides for optimal solar light capture. *Energy & Environmental Science*, 5(2):5814–5819, 2012b.

Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, et al. Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis*, 11(10):6059–6072, 2021.

Ziming Chen, Robert LZ Hoye, Hin-Lap Yip, Nadesh Fiuza-Maneiro, Iago López-Fernández, Clara Otero-Martínez, Lakshminarayana Polavarapu, Navendu Mondal, Alessandro Mirabelli, Miguel Anaya, et al. Roadmap on perovskite light-emitting diodes. *Journal of Physics: Photonics*, 6(3): 032501, 2024.

Jiucheng Cheng, Chunkai Zhang, and Lifeng Dong. A geometric-information-enhanced crystal graph network for predicting properties of materials. *Communications Materials*, 2(1):92, 2021.

Stefan Chmiela, Alexandre Tkatchenko, Huziel E Sauceda, Igor Poltavsky, Kristof T Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science Advances*, 3(5):e1603015, 2017.

Stefan Chmiela, Valentin Vassilev-Galindo, Oliver T Unke, Adil Kabylda, Huziel E Sauceda, Alexandre Tkatchenko, and Klaus-Robert Müller. Accurate global machine learning force fields for molecules with hundreds of atoms. *Science Advances*, 9(2):eadf0873, 2023.

Aron J Cohen, Paula Mori-Sánchez, and Weitao Yang. Insights into current limitations of density functional theory. *Science*, 321(5890):792–794, 2008.

Benjamin Coors, Alexandru Paul Condurache, and Andreas Geiger. Spherenet: Learning spherical representations for detection and classification in omnidirectional images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 518–533, 2018.

Jian S Dai. Euler–rodrigues formula variations, quaternion conjugation and intrinsic connections. *Mechanism and Machine Theory*, 92:144–152, 2015.

Kishalay Das, Pawan Goyal, Seung-Cheol Lee, Satadeep Bhattacharjee, and Niloy Ganguly. Crysmmnet: multimodal representation for crystal property prediction. In *Uncertainty in Artificial Intelligence*, pp. 507–517. PMLR, 2023.

Murray S Daw and Michael I Baskes. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Physical Review B*, 29(12):6443, 1984.

Fabio Di Fonzo, Carlo Spartaco Casari, Valeria Russo, Maria Francesca Brunella, A Li Bassi, and Carlo Enrico Bottani. Hierarchically organized nanostructuredtio2 for photocatalysis applications. *Nanotechnology*, 20(1):015604, 2008.

Alexandre Agm Duval, Victor Schmidt, Alex Hernández-Garcıa, Santiago Miret, Fragkiskos D Malliaros, Yoshua Bengio, and David Rolnick. Faenet: Frame averaging equivariant gnn for materials modeling. In *International Conference on Machine Learning*, pp. 9013–9033. PMLR, 2023.

Marcus Elstner and Gotthard Seifert. Density functional tight binding. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2011):20120483, 2014.

Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.

L. W. Finger and R. M. Hazen. Crystal structure and isothermal compression of fe_2o_3, cr_2o_3, and v_2o_3 to 50 kbars. *Journal of Applied Physics*, 51:5362–5367, 1980. doi: 10.1063/1.327451.

Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems*, 33:1970–1981, 2020.

Johannes Gasteiger, Shankari Giri, Johannes T Margraf, and Stephan Günnemann. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115*, 2020.

R. Grau-Crespo and R. Lopez-Cordero. Mos_2 structural properties. *Phys. Chem. Chem. Phys.*, 4: 4078, 2002.

M. Horn, C. R. Meagher, et al. Structure of anatase tio_2. *Zeitschrift für Kristallographie*, 136:273, 1972.

Anubhav Jain, Joseph Montoya, Shyam Dwaraknath, Nils ER Zimmermann, John Dagdelen, Matthew Horton, Patrick Huck, Donny Winston, Shreyas Cholia, Shyue Ping Ong, et al. The materials project: Accelerating materials design through theory-driven data and tools. *Handbook of Materials Modeling: Methods: Theory and Modeling*, pp. 1751–1784, 2020.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. *Advances in Neural Information Processing Systems*, 36:17464–17497, 2023.

Sarah John. Lattice functions, wavelet aliasing, and so (3) mappings of orthonormal filters. *Journal of Mathematical Physics*, 39(1):569–593, 1998.

Chaitanya K Joshi, Xiang Fu, Yi-Lun Liao, Vahe Gharakhanyan, Benjamin Kurt Miller, Anuroop Sriram, and Zachary W Ulissi. All-atom diffusion transformers: Unified generative modelling of molecules and materials. *arXiv preprint arXiv:2503.03965*, 2025.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Kuzma Khrabrov, Ilya Shenbin, Alexander Ryabov, Artem Tsypin, Alexander Telepov, Anton Alekseev, Alexander Grishin, Pavel Strashnov, Petr Zhilyaev, Sergey Nikolenko, et al. nabladft: Large-scale conformational energy and hamiltonian prediction benchmark and dataset. *Physical Chemistry Chemical Physics*, 24(42):25853–25863, 2022.

Byung-Hyun Kim, Jolla Kullgren, Matthew J Wolf, Kersti Hermansson, and Peter Broqvist. Multiscale modeling of agglomerated ceria nanoparticles: interface stability and oxygen vacancy formation. *Frontiers in Chemistry*, 7:203, 2019.

Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, et al. Pubchem 2025 update. *Nucleic Acids Research*, 53 (D1):D1516–D1525, 2025.

H. W. King. *CRC Handbook of Chemistry and Physics*. CRC Press, 83 edition, 2002a. Standard phase data for gold (Au).

H. W. King. *CRC Handbook of Chemistry and Physics*. CRC Press, 83 edition, 2002b. Standard phase data for silver (Ag).

Charles Kittel and Paul McEuen. *Introduction to solid state physics*. John Wiley & Sons, 2018.

Mustafa Kurban, Can Polat, Erchin Serpedin, and Hasan Kurban. Enhancing the electronic properties of tio2 nanoparticles through carbon doping: An integrated dftb and computer vision approach. *Computational Materials Science*, 244:113248, 2024.

Andrea C Levi and Miroslav Kotrla. Theory and simulation of crystal growth. *Journal of Physics: Condensed Matter*, 9(2):299, 1997.

Rongjin Li, Xiaotao Zhang, Huanli Dong, Qikai Li, Zhigang Shuai, and Wenping Hu. Gibbs–curie–wulff theorem in organic materials: a case study on the relationship between surface energy and crystal growth. *Advanced Materials*, 28(8):1697–1702, 2016.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022.

Hongsheng Liu, Gotthard Seifert, and Cristiana Di Valentin. An efficient way to model complex magnetite: Assessment of scc-dftb against dft. *The Journal of chemical physics*, 150(9), 2019.

Xiaoshan Luo, Zhenyu Wang, Qingchang Wang, Jian Lv, Lei Wang, Yanchao Wang, and Yanming Ma. Crystalflow: A flow-based generative model for crystalline materials. *arXiv preprint arXiv:2412.11693*, 2024.

Youzhi Luo, Keqiang Yan, and Shuiwang Ji. Graphdf: A discrete flow model for molecular graph generation. In *International Conference on Machine Learning*, pp. 7192–7203. PMLR, 2021.

David W Lyons. An elementary introduction to the hopf fibration. *Mathematics Magazine*, 76(2): 87–98, 2003.

Avik Mahata, Tanmoy Mukhopadhyay, and Mohsen Asle Zaeem. Modified embedded-atom method interatomic potentials for al-cu, al-fe and al-ni binary alloys: From room temperature to melting point. *Computational Materials Science*, 201:110902, 2022.

Alessandro Mattoni, Alessio Filippetti, and Claudia Caddeo. Modeling hybrid perovskites by molecular dynamics. *Journal of Physics: Condensed Matter*, 29(4):043001, 2016.

Benjamin Kurt Miller, Ricky TQ Chen, Anuroop Sriram, and Brandon M Wood. Flowmm: Generating materials with riemannian flow matching. In *Forty-first International Conference on Machine Learning*, 2024.

R. H. Mitchell and M. A. Carpenter. Physics and chemistry of $srtio_3$ perovskites. *Physics and Chemistry of Minerals*, 27:583, 2000.

Svetozar Najman, Hsin-An Chen, Hsin-Yi Tiffany Chen, and Chun-Wei Pao. Surface structures and equilibrium shapes of layered 2d ruddlesden-popper perovskite crystals from density functional theory calculations. *Materials Today Communications*, 26:101745, 2021.

Maylis Orio, Dimitrios A Pantazis, and Frank Neese. Density functional theory. *Photosynthesis Research*, 102:443–453, 2009.

A Paszke. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.

Anyang Peng, Chun Cai, Mingyu Guo, Duo Zhang, Chengqian Zhang, Antoine Loew, Linfeng Zhang, and Han Wang. Lambench: A benchmark for large atomic models. *arXiv preprint arXiv:2504.19578*, 2025.

Chris J Pickard. Airss data for carbon at 10gpa and the c+ n+ h+ o system at 1gpa. *(No Title)*, 2020.

Can Polat, Mustafa Kurban, and Hasan Kurban. Multimodal neural network-based predictive modeling of nanoparticle properties from pure compounds. *Machine Learning: Science and Technology*, 5(4):045062, 2024.

Can Polat, Erchin Serpedin, Mustafa Kurban, and Hasan Kurban. Crysmtm: a multiphase, temperature-resolved, multimodal dataset for crystalline materials. *Machine Learning: Science and Technology*, 6(3):030603, 2025.

Tingting Qi, Charles W Bauschlicher Jr, John W Lawson, Tapan G Desai, and Evan J Reed. Comparison of reaxff, dftb, and dft for phenolic pyrolysis. 1. molecular dynamics simulations. *The Journal of Physical Chemistry A*, 117(44):11115–11125, 2013.

Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):1–7, 2014.

Emilie Ringe, Richard P Van Duyne, and Laurence D Marks. Kinetic and thermodynamic modified wulff constructions for twinned nanoparticles. *The Journal of Physical Chemistry C*, 117(31): 15859–15870, 2013.

Zachary A Rollins, Alan C Cheng, and Essam Metwally. Molprop: Molecular property prediction with multimodal language and graph fusion. *Journal of Cheminformatics*, 16(1):56, 2024.

William J Rucklidge. Efficiently locating objects using the hausdorff distance. *International Journal of computer vision*, 24:251–270, 1997.

Lars Ruddigkeit, Ruud Van Deursen, Lorenz C Blum, and Jean-Louis Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17. *Journal of Chemical Information and Modeling*, 52(11):2864–2875, 2012.

M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical Review Letters*, 108:058301, 2012.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International Conference on Machine Learning*, pp. 9323–9332. PMLR, 2021.

Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.

Kristof T Schütt, Huziel E Sauceda, P-J Kindermans, Alexandre Tkatchenko, and K-R Müller. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24), 2018.

Arkadiy Simonov and Andrew L Goodwin. Designing disorder into crystalline materials. *Nature Reviews Chemistry*, 4(12):657–673, 2020.

Chris-Kriton Skylaris, Peter D Haynes, Arash A Mostofi, and Mike C Payne. Introducing onetep: Linear-scaling density functional simulations on parallel computers. *The Journal of Chemical Physics*, 122(8), 2005.

Fernand Spiegelman, Nathalie Tarrat, Jérôme Cuny, Leo Dontot, Evgeny Posenitskiy, Carles Martí, Aude Simon, and Mathias Rapacioli. Density-functional tight-binding: basic concepts and applications to molecules and clusters. *Advances in Physics: X*, 5(1):1710252, 2020.

Anuroop Sriram, Benjamin Miller, Ricky TQ Chen, and Brandon Wood. Flowllm: Flow matching for material generation with large language models as base distributions. *Advances in Neural Information Processing Systems*, 37:46025–46046, 2024.

Richard P Stanley. The fibonacci lattice. *The Fibonacci Quarterly*, 13(3):215–232, 1975.

Thomas Surek. Crystal growth and materials research in photovoltaics: progress and challenges. *Journal of Crystal growth*, 275(1-2):292–304, 2005.

Philipp Thölke and Gianni De Fabritiis. Torchmd-net: equivariant transformers for neural network based molecular potentials. *arXiv preprint arXiv:2202.02541*, 2022.

Richard Tran, Janice Lan, Muhammed Shuaibi, Brandon M Wood, Siddharth Goyal, Abhishek Das, Javier Heras-Domingo, Adeesh Kolluru, Ammar Rizvi, Nima Shoghi, et al. The open catalyst 2022 (oc22) dataset and challenges for oxide electrocatalysts. *ACS Catalysis*, 13(5):3066–3084, 2023.

Aron Walsh, elds22, Federico Brivio, and Jarvist Moore Frost. Wmd-group/hybrid-perovskites: Collection 1 (v1.0). `https://doi.org/10.5281/zenodo.2641358`, 2019. Hybrid perovskite $CH_3NH_3PbI_3$ structural data.

R. W. G. Wyckoff. *Crystal Structures Volume 1*. Interscience Publishers, 1963a.

R. W. G. Wyckoff. *Crystal Structures Volume 1*. Interscience Publishers, 1963b.

R. W. G. Wyckoff. *Crystal Structures Volume 1*. Interscience Publishers, 1963c.

Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197*, 2021.

Ruo Xi Yang, Caitlin A McCandler, Oxana Andriuc, Martin Siron, Rachel Woods-Robinson, Matthew K Horton, and Kristin A Persson. Big data in a nano world: a review on computational, data-driven design of nanomaterials structures, properties, and synthesis. *ACS Nano*, 16 (12):19873–19891, 2022.

Sherry Yang, KwangHwan Cho, Amil Merchant, Pieter Abbeel, Dale Schuurmans, Igor Mordatch, and Ekin Dogus Cubuk. Scalable diffusion for materials generation. *arXiv preprint arXiv:2311.09235*, 2023.

Anna Yershova, Swati Jain, Steven M Lavalle, and Julie C Mitchell. Generating uniform incremental grids on so (3) using the hopf fibration. *The International Journal of Robotics Research*, 29(7): 801–812, 2010.

Haiyang Yu, Meng Liu, Youzhi Luo, Alex Strasser, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Qh9: A quantum hamiltonian prediction benchmark for qm9 molecules. *Advances in Neural Information Processing Systems*, 36, 2024.

Haoyu S Yu, Shaohong L Li, and Donald G Truhlar. Perspective: Kohn-sham density functional theory descending a staircase. *The Journal of Chemical Physics*, 145(13), 2016.

Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Sasha Shysheya, Jonathan Crabbé, Lixin Sun, Jake Smith, et al. Mattergen: a generative model for inorganic materials design. *arXiv preprint arXiv:2312.03687*, 2023.

Guishan Zheng, Stephan Irle, and Keiji Morokuma. Performance of the dftb method in comparison to dft and semiempirical methods for geometries and energies of c20–c86 fullerene isomers. *Chemical Physics Letters*, 412(1-3):210–216, 2005.