
Crystal Diffusion Variational Autoencoder for Periodic Material Generation

Tian Xie,* Xiang Fu,* Octavian-Eugen Ganea,* Regina Barzilay, & Tommi Jaakkola

Computer Science and Artificial Intelligence Laboratory

Massachusetts Institute of Technology

Cambridge, MA 02139, USA

{txie, xiangfu, oct, regina, tommi}@csail.mit.edu

Abstract

Generating the periodic structure of stable materials is a long-standing challenge for the material design community. This task is difficult because stable materials only exist in a low-dimensional subspace of all possible periodic arrangements of atoms: 1) the coordinates must lie in the local energy minimum defined by quantum mechanics, and 2) global stability also requires the structure to follow the complex, yet specific bonding preferences between different atom types. Existing methods fail to incorporate these factors and often lack proper invariances. We propose a Crystal Diffusion Variational Autoencoder (CDVAE) that captures the physical inductive bias of material stability. By learning from the data distribution of stable materials, the decoder generates materials in a diffusion process that moves atomic coordinates towards a lower energy state and updates atom types to satisfy bonding preferences between neighbors. Our model also explicitly encodes interactions across periodic boundaries and respects permutation, translation, rotation, and periodic invariances. We significantly outperform past methods in three tasks: 1) reconstructing the input structure, 2) generating valid, diverse, and realistic materials, and 3) generating materials that optimize a specific property. We also provide several standard datasets and evaluation metrics for the broader machine learning community.

1 Introduction

Solid state materials, represented by the periodic arrangement of atoms in the 3D space, are the foundation of many key technologies including solar cells, batteries, and catalysis [1]. Despite the rapid progress of molecular generative models and their significant impact on drug discovery, the problem of material generation has many unique challenges. Compared with small molecules, materials have more complex periodic 3D structures and cannot be adequately represented by a simple graph like small molecules (Figure 1). In addition, materials can be made up of more than 100 elements in the periodic table, while molecules are generally only made up of a small subset of atoms such as carbon, oxygen, and hydrogen. Finally, the data for training ML models for material design is limited. There are only $\sim 200k$ experimentally known inorganic materials, collected by the ICSD [2], in contrast to close to a billion purchasable molecules in ZINC [3].

The key challenge of this task is in generating *stable* materials. Such materials only exist in a low-dimensional subspace of all possible periodic arrangements of atoms: 1) the atom coordinates must lie in the local energy minimum defined by quantum mechanics (QM); 2) global stability also requires the structure to follow the complex, yet specific bonding preferences between different atom types (section 2). The issue of stability is unique to material generation because valency checkers assessing molecular stability are not applicable to materials. Moreover, we also have to encode the

*Equal contribution. Correspondence to: Tian Xie at txie@csail.mit.edu

interactions crossing periodic boundaries (Figure 1, middle), and satisfy permutation, translation, rotation, and periodic invariances (Appendix C). Our goal is to learn representations that can learn features of stable materials from data, while adhering to the above invariance properties.

We address these challenges by learning a variational autoencoder (VAE) [5] to generate stable 3D materials directly from a latent representation without intermediates like graphs. The key insight is to exploit the fact that all materials in the data distribution are stable, therefore if noise is added to the ground truth structure, denoising it back to its original structure will likely increase stability. We capture this insight by designing a noise conditional score network (NCSN) [6] as our decoder: 1) the decoder outputs gradients that drive the atom coordinates to the energy local minimum; 2) it also updates atom types based on the neighbors to capture the specific local bonding preferences (e.g., Si-O is preferred over Si-Si and O-O in SiO₂). During generation, materials are generated using Langevin dynamics that gradually deforms an initial random structure to a stable structure. To capture the necessary invariances and encode the interactions crossing periodic boundaries, we use SE(3) equivariant graph neural networks adapted with periodicity (PGNNs) for both the encoder and decoder of our VAE.

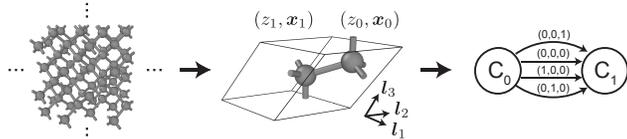


Figure 1: The periodic structure of diamond. The left shows the infinite periodic structure, the middle shows a unit cell representing the periodic structure, and the right shows a multi-graph [4] representation.

Our theoretical analysis (Appendix E) further reveals an intriguing connection between the gradient field learned by our decoder and an harmonic force field. De facto, the decoder utilizes the latter to estimate the forces on atoms when their coordinates deviate from the equilibrium positions. Consequently, this formulation provides an important physical inductive bias for generating stable materials.

In this work, we propose Crystal Diffusion Variational AutoEncoder (CDVAE) to generate stable materials by learning from the data distribution of known materials. Our main contributions include:

- We curate 3 standard datasets from QM simulations and create a set of physically meaningful tasks and metrics for the problem of material generation.
- We incorporate stability as an inductive bias by designing a noise conditional score network as the decoder of our VAE, which allows us to generate significantly more realistic materials.
- We encode permutation, translation, rotation, and periodic invariances, as well as interactions crossing periodic boundaries with SE(3) equivariant GNNs adapted with periodicity.
- Empirically, our model significantly outperforms past methods in tasks including reconstructing an input structure, generating valid, diverse, and realistic materials, and generating materials that optimize specific properties.

2 Preliminaries

Periodic structure of materials. Any material structure can be represented as the periodic arrangement of atoms in the 3D space. As illustrated in Figure 1, we can always find a repeating unit, i.e. a unit cell, to describe the infinite periodic structure of a material. A unit cell that includes N atoms can be fully described by 3 sets: 1) atom types $\mathbf{A} = (a_0, \dots, a_N) \in \mathbb{A}^N$, where \mathbb{A} denotes the set of all chemical elements; 2) atom coordinates $\mathbf{X} = (\mathbf{x}_0, \dots, \mathbf{x}_N) \in \mathbb{R}^{N \times 3}$; and 3) periodic lattice $\mathbf{L} = (l_1, l_2, l_3) \in \mathbb{R}^{3 \times 3}$. The periodic lattice defines the periodic translation symmetry of the material. Given $\mathbf{M} = (\mathbf{A}, \mathbf{X}, \mathbf{L})$, the infinite periodic structure can be represented as,

$$\{(a'_i, \mathbf{x}'_i) | a'_i = a_i, \mathbf{x}'_i = \mathbf{x}_i + k_1 l_1 + k_2 l_2 + k_3 l_3, k_1, k_2, k_3 \in \mathbb{Z}\}, \quad (1)$$

where k_1, k_2, k_3 are any integers that translate the unit cell using \mathbf{L} to tile the entire 3D space.

The composition of a material denotes the ratio of different elements that the material is composed of. Given the atom types of a material with N atoms $\mathbf{A} \in \mathbb{A}^N$, the composition can be represented as $c \in \mathbb{R}^{|\mathbb{A}|}$, where $c_i > 0$ denotes the percentage of atom type i and $\sum_i c_i = 1$. For example, the composition of diamond in Figure 1 has $c_6 = 1$ and $c_i = 0$ for $i \neq 6$ because 6 is the atomic number of carbon.

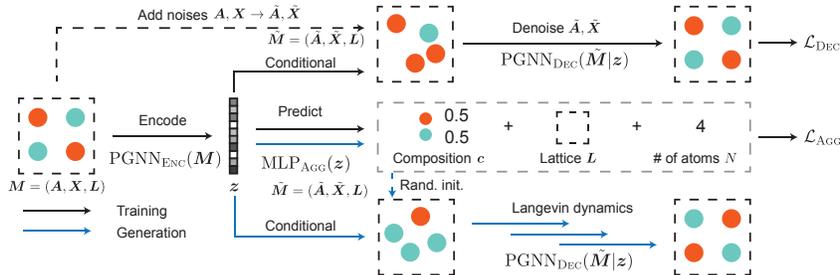


Figure 2: Overview of the proposed CDVAE approach.

Problem definition and its physical origin. Our goal is to generate novel, stable materials $M = (A, X, L) \in \mathbb{A}^N \times \mathbb{R}^{N \times 3} \times \mathbb{R}^{3 \times 3}$ by learning from the empirical distribution of experimentally observed stable materials. The space of stable materials is a subspace in $\mathbb{A}^N \times \mathbb{R}^{N \times 3} \times \mathbb{R}^{3 \times 3}$ that satisfies the following constraints. 1) The materials lie in the local minimum of the energy landscape defined by quantum mechanics, with respect to the atom coordinates and lattice, i.e. $\partial E / \partial X = 0$ and $\partial E / \partial L = 0$. 2) The material is globally stable and thus cannot decompose into nearby phases. Global stability is strongly related to bonding preferences between neighboring atoms. For example, in SiO_2 , each Si is surrounded by 4 O and each O is surrounded by 2 Si. This configuration is caused by the stronger bonding preferences between Si-O than Si-Si and O-O.

3 Overview of Proposed Method

Our approach generates new materials via a two-step process: 1) We sample a z from the latent space and use it to predict 3 aggregated properties of a material: composition (c), lattice (L), and number of atoms (N), which are then used to randomly initialize a material structure $\tilde{M} = (\tilde{A}, \tilde{X}, L)$. 2) We perform Langevin dynamics to simultaneously denoise \tilde{X} and \tilde{A} conditioned on z to improve both the local and global stability of \tilde{M} and generate the final structure of the new material.

To train our model, we optimize 3 networks concurrently using stable materials $M = (A, X, L)$ sampled from the data distribution. 1) A periodic GNN encoder $\text{PGNN}_{\text{ENC}}(M)$ that encodes M into a latent representation z . 2) A property predictor $\text{MLP}_{\text{AGG}}(z)$ that predicts the c , L , and N of M from z . 3) A periodic GNN decoder $\text{PGNN}_{\text{DEC}}(\tilde{M}|z)$ that denoises both \tilde{X} and \tilde{A} conditioned on z . For 3), the noisy structure $\tilde{M} = (\tilde{A}, \tilde{X}, L)$ is obtained by adding different levels of noise to X and A . The noise schedules are defined by the predicted aggregated properties, with the motivation of simplifying the task for our decoder from denoising an arbitrary random structure from over ~ 100 elements to a constrained random structure from predicted properties. We train all three networks together by minimizing a combined loss including the aggregated property loss \mathcal{L}_{AGG} , decoder denoising loss \mathcal{L}_{DEC} , and a KL divergence loss \mathcal{L}_{KL} for the VAE.

To capture the interactions across periodic boundaries, we employ a multi-graph representation (Appendix B) for both M and \tilde{M} . We also use SE(3) equivariant GNNs adapted with periodicity as both the encoder and the decoder to ensure the permutation, translation, rotation, and periodic invariances of our model. The CDVAE is summarized in Figure 2 and we explain the individual components of our method in Appendix D and the implementation details in Appendix F.

4 Experiments

We evaluate multiple aspects of material generation that are related to real-world material discovery process. Past studies in this field used very different tasks and metrics, making them difficult to compare. Building upon past studies [7, 8], we create a set of standard tasks, datasets, and metrics to evaluate and compare models for material generation. Models are trained on single RTX 2080 GPUs. Training generally takes 5-10 hours, and the Langevin dynamics takes 1-3 hours.

Tasks. We focus on 2 tasks for material generation. 1) **Reconstruction** evaluates the ability of the model to reconstruct the original material from its latent representation z . 2) **Generation** evaluates the validity, coverage, and property statistics of material structures generated by the model. 3)

Table 1: Reconstruction performance.

Method	Match rate (%) \uparrow			RMSE \downarrow		
	Perov-5	Carbon-24	MP-20	Perov-5	Carbon-24	MP-20
FTCP	99.34	62.28	69.89	0.0259	0.2563	0.1593
Cond-DFC-VAE	51.65	–	–	0.0217	–	–
CDVAE	97.52	55.22	45.43	0.0156	0.1251	0.0356

Property optimization evaluates the model’s ability to generate materials that are optimized for a specific property.

Datasets. We curated 3 datasets obtained from quantum mechanical simulations. 1) **Perov-5** [9, 10] includes 18928 perovskite materials that share the same structure but differ in composition. 2) **Carbon-24** [11] includes 10153 materials that are all made up with carbon atoms but differ in structures. 3) **MP-20** [12] includes 45231 materials that differ in both structure and composition. Perov-5 and Carbon-24 are toy datasets and many materials are hypothetical. MP-20 is a realistic dataset and includes most experimentally known materials with no more than 20 atoms. We use a 60-20-20 random split for all of our experiments. Details regarding dataset curation can be found at Appendix A.

Baselines. We compare CDVAE with the following 4 baselines, which include the latest coordinate-based, voxel-based, and 3D molecule generation methods. **FTCP** [8] is a crystal representation that concatenates real-space properties (atom positions, atom types, etc.) and Fourier-transformed momentum-space properties (diffraction pattern). A 1D CNN-VAE is trained over this representation for crystal generation. **Cond-DFC-VAE** [7] encodes and generates crystals with 3D density maps, while employing several modifications over the previous Voxel-VAE [13] method. However, the effectiveness is only demonstrated for cubic systems, limiting its usage to the Perov-5 dataset. **G-SchNet** [14] is an auto-regressive model that generates 3D molecules by performing atom-by-atom completion using SchNet [15]. Since G-SchNet is unaware of periodicity and cannot generate the lattice L . We adapt G-SchNet to our material generation tasks by constructing the smallest oriented bounding box with PCA such that the introduced periodicity does not cause structural invalidity. **P-G-SchNet** is our modified G-SchNet that incorporates periodicity. During training, the SchNet encoder inputs the partial periodic structure to predict next atoms. During generation, we first randomly sample a lattice L from training data and autoregressively generate the periodic structure.

4.1 Material reconstruction

Setup. The first task is to reconstruct the material from its latent representation. We evaluate reconstruction performance by matching the generated structure and the input structure for all materials in the test set. Definitions of match rate and RMSE are detailed in Appendix G.1.

Results. The reconstructed structures are shown in Figure 3 and the metrics are in Table 1. Since our model is SE(3) invariant, the generated structures may be a translated (or rotated) version of the ground truth structure. Our model has a lower RMSE than all other models, indicating its stronger capability to generate realistic structures. FTCP has a higher match rate than our model in Perov-5 and MP-20. This could be explained by fact that the same set of local structures can be assembled into different materials globally (e.g., two different crystal forms of ZnS). Our model is SE(3) invariant and only encodes local structures, while FTCP directly encodes the global structures via absolute coordinates and types of each atom. In Figure 4, we show that CDVAE can generate different plausible arrangement of atoms by sampling 3 Langevin dynamics with different random seeds from the same z .

4.2 Material generation

Setup. The second task is to generate novel, stable materials that are distributionally similar to the test materials. The only high-fidelity evaluation of stability of generated materials is to perform

²Some metrics unsuitable for specific datasets have “–” values in the table.

³For comparison, the ground truth structure validity is 100.0 % for all datasets, and ground truth composition validity is 98.60 %, 100.0 %, 91.13 % for Perov-5, Carbon-24, and MP-20.

⁴Due to the low validity of FTCP, we instead randomly generate 100,000 materials from $\mathcal{N}(0, 1)$ and use 1,000 materials from those valid ones to compute diversity and property statistics metrics.

Table 2: Generation performance².

Method	Data	Validity (%) ³ ↑		COV (%) ↑		Property Statistics ↓		
		Struc.	Comp.	R.	P.	ρ	E	# elem.
FTCP ⁴	Perov-5	0.24	54.24	0.00	0.00	10.27	156.0	0.6297
	Carbon-24	0.08	–	0.00	0.00	5.206	19.05	–
	MP-20	1.55	48.37	4.72	0.09	23.71	160.9	0.7363
Cond-DFC-VAE	Perov-5	73.60	82.95	73.92	10.13	2.268	4.111	0.8373
	G-SchNet	99.92	98.79	0.18	0.23	1.625	4.746	0.03684
P-G-SchNet	Carbon-24	99.94	–	0.00	0.00	0.9427	1.320	–
	MP-20	99.65	75.96	38.33	99.57	3.034	42.09	0.6411
	Perov-5	79.63	99.13	0.37	0.25	0.2755	1.388	0.4552
	Carbon-24	48.39	–	0.00	0.00	1.533	134.7	–
CDVAE	MP-20	77.51	76.40	41.93	99.74	4.04	2.448	0.6234
	Perov-5	100.0	98.59	99.45	98.46	0.1258	0.0264	0.0628
	Carbon-24	100.0	–	99.80	83.08	0.1407	0.2850	–
	MP-20	100.0	86.70	99.15	99.49	0.6875	0.2778	1.432

Table 3: Property optimization performance.

Method	Perov-5			Carbon-24			MP-20		
	SR5	SR10	SR15	SR5	SR10	SR15	SR5	SR10	SR15
FTCP	0.06	0.11	0.16	0.0	0.0	0.0	0.02	0.04	0.05
Cond-DFC-VAE	0.55	0.64	0.69	–	–	–	–	–	–
CDVAE	0.52	0.65	0.79	0.0	0.06	0.06	0.78	0.86	0.90

QM calculations, but it is computationally prohibitive to use QM for computing evaluation metrics. We developed several physically meaningful metrics to evaluate the validity, coverage (COV), and property statistics of generated materials, as detailed in Appendix G.2. Validity and coverage are computed over 10,000 materials randomly sampled from $\mathcal{N}(0, 1)$. Property statistics is computed over 1,000 valid materials randomly sampled from those that pass the validity test.

Results. The generated structures are shown in Figure 5 and the metrics are in Table 2. Our model achieves a higher validity than FTCP, Cond-DFC-VAE, and P-G-SchNet, while G-SchNet achieves a similar validity as ours. The lower structural validity in P-G-SchNet than G-SchNet is likely due to the difficulty of avoiding atom collisions during the autoregressive generation inside a finite periodic box. On the contrary, our G-SchNet baseline constructs the lattice box after the 3D positions of all atoms are generated, and the construction explicitly avoids introducing invalidity. Furthermore, our model also achieves higher COV-R and COV-P than all other models, except in MP-20 our COV-P is similar to G-SchNet and P-G-SchNet. These results indicate that our model generates both diverse (COV-R) and high quality (COV-P) materials. Finally, our model also significantly outperforms all other models in the property statistics of density and energy, further confirming the high quality of generated materials. We observe that our method tends to generate more elements in a material than ground truth, which explains the lower performance in the statistics of # of elems. than G-SchNet. We hypothesize this is due to the non-Gaussian statistical structure of ground truth materials, and using a more complex prior, e.g., a flow-model-transformed Gaussian [16], might resolve this issue.

4.3 Property optimization

Setup. The third task is to generate materials that optimize a specific property, with detailed methods in Appendix G.3. For all methods, we generate 100 materials following the protocol above. We use the independent property predictor to predict the properties for evaluation. We report the success rate (SR) as the percentage of materials achieving 5, 10, and 15 percentiles of the target property distribution. Our task is to minimize formation energy per atom for all 3 datasets.

Results. The performance is shown in Table 3. We significantly outperform FTCP, while having a similar performance as Cond-DFC-VAE in Perov-5 (Cond-DFC-VAE cannot work for Carbon-24 and MP-20). Both G-SchNet and P-G-SchNet are incapable of property optimization⁵. We note that all models perform poorly on the Carbon-24 dataset, which might be explained by the complex and diverse 3D structures of carbon.

⁵Very recently the authors published an improved version for conditional generation [17] but the code has not been released yet.

5 Broader Impact

The ability to generate novel, stable materials has the potential to accelerate the discovery of materials and become a valuable tool for tackling societal challenges like climate change. The potential negative effects include the possibility of disrupting existing industries and the method being used for producing harmful materials.

Acknowledgments and Disclosure of Funding

We thank Peter Mikhael, Jason Yim, Rachel Wu, Bracha Laufer, Gabriele Corso, Felix Faltings, Bowen Jing, and the rest of the RB and TJ group members for their helpful comments and suggestions. The authors gratefully thank DARPA (HR00111920025) and the consortium Machine Learning for Pharmaceutical Discovery and Synthesis (mlpds.mit.edu) for support.

References

- [1] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," *Nature*, vol. 559, no. 7715, pp. 547–555, 2018.
- [2] A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch, "New developments in the inorganic crystal structure database (icsd): accessibility in support of materials research and design," *Acta Crystallographica Section B: Structural Science*, vol. 58, no. 3, pp. 364–369, 2002.
- [3] J. J. Irwin and B. K. Shoichet, "Zinc- a free database of commercially available compounds for virtual screening," *Journal of chemical information and modeling*, vol. 45, no. 1, pp. 177–182, 2005.
- [4] T. Xie and J. C. Grossman, "Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties," *Physical review letters*, vol. 120, no. 14, p. 145301, 2018.
- [5] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2014.
- [6] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada* (H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, eds.), pp. 11895–11907, 2019.
- [7] C. J. Court, B. Yildirim, A. Jain, and J. M. Cole, "3-d inorganic crystal structure generation and property prediction via representation learning," *Journal of chemical information and modeling*, vol. 60, no. 10, pp. 4518–4535, 2020.
- [8] Z. Ren, J. Noh, S. Tian, F. Oviedo, G. Xing, Q. Liang, A. Aberle, Y. Liu, Q. Li, S. Jayavelu, *et al.*, "Inverse design of crystals using generalized invertible crystallographic representation," *arXiv preprint arXiv:2005.07609*, 2020.
- [9] I. E. Castelli, D. D. Landis, K. S. Thygesen, S. Dahl, I. Chorkendorff, T. F. Jaramillo, and K. W. Jacobsen, "New cubic perovskites for one-and two-photon water splitting using the computational materials repository," *Energy & Environmental Science*, vol. 5, no. 10, pp. 9034–9043, 2012.
- [10] I. E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen, and K. W. Jacobsen, "Computational screening of perovskite metal oxides for optimal solar light capture," *Energy & Environmental Science*, vol. 5, no. 2, pp. 5814–5819, 2012.
- [11] C. J. Pickard, "Airss data for carbon at 10gpa and the c+n+h+o system at 1gpa," 2020.
- [12] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, *et al.*, "Commentary: The materials project: A materials genome approach to accelerating materials innovation," *APL materials*, vol. 1, no. 1, p. 011002, 2013.

- [13] J. Hoffmann, L. Maestrati, Y. Sawada, J. Tang, J. M. Sellier, and Y. Bengio, "Data-driven approach to encoding and decoding 3-d crystal structures," *arXiv preprint arXiv:1909.00949*, 2019.
- [14] N. Gebauer, M. Gastegger, and K. Schütt, "Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules," in *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), vol. 32, Curran Associates, Inc., 2019.
- [15] K. T. Schütt, H. E. Saucedo, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SchNet—a deep learning architecture for molecules and materials," *The Journal of Chemical Physics*, vol. 148, no. 24, p. 241722, 2018.
- [16] G. Yang, X. Huang, Z. Hao, M.-Y. Liu, S. Belongie, and B. Hariharan, "Pointflow: 3d point cloud generation with continuous normalizing flows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4541–4550, 2019.
- [17] N. W. Gebauer, M. Gastegger, S. S. Hessmann, K.-R. Müller, and K. T. Schütt, "Inverse design of 3d molecular structures with conditional generative neural networks," *arXiv preprint arXiv:2109.04824*, 2021.
- [18] C. J. Pickard and R. Needs, "High-pressure phases of silane," *Physical review letters*, vol. 97, no. 4, p. 045504, 2006.
- [19] C. J. Pickard and R. Needs, "Ab initio random structure searching," *Journal of Physics: Condensed Matter*, vol. 23, no. 5, p. 053201, 2011.
- [20] A. F. Wells *et al.*, *Three dimensional nets and polyhedra*. Wiley, 1977.
- [21] M. O’Keeffe and B. G. Hyde, "Plane nets in crystal chemistry," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 295, no. 1417, pp. 553–618, 1980.
- [22] H. Pan, A. M. Ganose, M. Horton, M. Aykol, K. A. Persson, N. E. Zimmermann, and A. Jain, "Benchmarking coordination number prediction algorithms on inorganic crystal structures," *Inorganic chemistry*, vol. 60, no. 3, pp. 1590–1603, 2021.
- [23] R. W. Grosse-Kunstleve, N. K. Sauter, and P. D. Adams, "Numerically stable algorithms for the computation of reduced unit cells," *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 60, no. 1, pp. 1–6, 2004.
- [24] C. Shi, S. Luo, M. Xu, and J. Tang, "Learning gradient fields for molecular conformation generation," in *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event* (M. Meila and T. Zhang, eds.), vol. 139 of *Proceedings of Machine Learning Research*, pp. 9558–9568, PMLR, 2021.
- [25] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, "Python materials genomics (pymatgen): A robust, open-source python library for materials analysis," *Computational Materials Science*, vol. 68, pp. 314–319, 2013.
- [26] J. Klicpera, S. Giri, J. T. Margraf, and S. Günnemann, "Fast and uncertainty-aware directional message passing for non-equilibrium molecules," *arXiv preprint arXiv:2011.14115*, 2020.
- [27] J. Klicpera, J. Groß, and S. Günnemann, "Directional message passing for molecular graphs," in *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, OpenReview.net, 2020.
- [28] J. Klicpera, F. Becker, and S. Günnemann, "Gemnet: Universal directional graph neural networks for molecules," *arXiv preprint arXiv:2106.08903*, 2021.
- [29] L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, *et al.*, "Open catalyst 2020 (oc20) dataset and community challenges," *ACS Catalysis*, vol. 11, no. 10, pp. 6059–6072, 2021.

- [30] D. W. Davies, K. T. Butler, A. J. Jackson, J. M. Skelton, K. Morita, and A. Walsh, “Smact: Semiconducting materials by analogy and chemical theory,” *Journal of Open Source Software*, vol. 4, no. 38, p. 1361, 2019.
- [31] M. Xu, S. Luo, Y. Bengio, J. Peng, and J. Tang, “Learning neural generative dynamics for molecular conformation generation,” in *International Conference on Learning Representations*, 2021.
- [32] O.-E. Ganea, L. Pattanaik, C. W. Coley, R. Barzilay, K. F. Jensen, W. H. Green, and T. S. Jaakkola, “Geomol: Torsional geometric generation of molecular 3d conformer ensembles,” *arXiv preprint arXiv:2106.07802*, 2021.
- [33] W. Jin, R. Barzilay, and T. Jaakkola, “Junction tree variational autoencoder for molecular graph generation,” in *International conference on machine learning*, pp. 2323–2332, PMLR, 2018.

Checklist

1. For all authors...
 - (a) Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope? [\[Yes\]](#)
 - (b) Did you describe the limitations of your work? [\[Yes\]](#)
 - (c) Did you discuss any potential negative societal impacts of your work? [\[Yes\]](#)
 - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [\[Yes\]](#)
2. If you are including theoretical results...
 - (a) Did you state the full set of assumptions of all theoretical results? [\[N/A\]](#)
 - (b) Did you include complete proofs of all theoretical results? [\[N/A\]](#)
3. If you ran experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [\[No\]](#) We are still in the process of making the code and data available to the public.
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [\[Yes\]](#)
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [\[No\]](#) The error bars with respect to the random seed is very small compared with the performance difference of the models.
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [\[Yes\]](#)
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
 - (a) If your work uses existing assets, did you cite the creators? [\[Yes\]](#)
 - (b) Did you mention the license of the assets? [\[Yes\]](#) We included licenses of Carbon-24 and MP-20. We cannot find the license of Perov-5.
 - (c) Did you include any new assets either in the supplemental material or as a URL? [\[Yes\]](#) We curated 3 datasets including Perov-5, Carbon-24, and MP-20 from public available data. We are still in the process of making them available.
 - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? [\[No\]](#) The original data is public available.
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [\[No\]](#)
5. If you used crowdsourcing or conducted research with human subjects...
 - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [\[N/A\]](#)
 - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [\[N/A\]](#)
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [\[N/A\]](#)

A Dataset curation

A.1 Perov-5

Perovskite is a class of materials that share a similar structure and have the general chemical formula ABX_3 . The ideal perovskites have a cubic structure, where the site A atom sits at a corner position, the site B atom sits at a body centered position and site X atoms sit at face centered positions. Perovskite materials are known for their wide applications and we curate the Perov-5 dataset from an open database that was originally developed for water splitting [9, 10].

All 18928 materials in the original database are included. In the database, A, B can be any non-radioactive metals and X can be one or several elements from O, N, S, and F. Note that there can be multiple different X atoms in the same material. All materials in Perov-5 are relaxed using density functional theory (DFT), and their relaxed structure can deviate significantly from the ideal structures. A significant portion of the materials are not thermodynamically stable, i.e. they will decompose nearby phases and cannot be synthesized.

A.2 Carbon-24

Carbon-24 includes various carbon structures obtained via *ab initio* random structure searching (AIRSS) [18, 19] performed at 10 GPa.

The original dataset includes 101529 carbon structures, and we selected the 10% of the carbon structure with the lowest energy per atom to create Carbon-24. All 10153 structures in Carbon-24 are relaxed using DFT. The most stable structure is diamond at 10 GPa. All rest structures are thermodynamically unstable but may be kinetically stable. Most of the structures cannot be synthesized.

The dataset is licensed under Attribution 4.0 International according to the original authors [11].

A.3 MP-20

MP-20 includes almost all experimentally stable materials from the Materials Project [12] with unit cells including at most 20 atoms. We only include materials that are originally from ICSD [2] to ensure the experimental stability, and these materials represent the majority of experimentally known materials with at most 20 atoms in unit cells.

To ensure stability, we only select materials with energy above the hull smaller than 0.08 eV/atom and formation energy smaller than 2 eV/atom, following [8]. Differing from [8], we do not constrain the number of unique elements per material. All materials in MP-20 are relaxed using DFT. Most materials are thermodynamically stable and have been synthesized.

The dataset is licensed under Attribution 4.0 International according to the original authors [12].

B Multi-graph representation for materials

Materials can be represented as a directed multi-graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ to encode the periodic structures following [4, 20, 21], where $\mathcal{V} = \{v_1, \dots, v_N\}$ is the set of nodes representing atoms and $\mathcal{E} = \{e_{ij, (k_1, k_2, k_3)} | i, j \in \{1, \dots, N\}, k_1, k_2, k_3 \in \mathbb{Z}\}$ is the set of edges representing bonds. $e_{ij, (k_1, k_2, k_3)}$ denotes a directed edge from node i at the original unit cell to node j at the cell translated by $k_1\mathbf{l}_1 + k_2\mathbf{l}_2 + k_3\mathbf{l}_3$ (in Figure 1, (k_1, k_2, k_3) are labeled on top of edges). For materials, there is no unique way to define edges (bonds) and the edges are often computed using k-nearest neighbor (KNN) approaches under periodicity, or using more advanced methods such as CrystalNN [22]. Given this directed multi-graph, popular message-passing neural networks and SE(3)-equivariant networks can be used for the representation learning of materials.

C Invariances for materials.

The structure of materials do not change under several invariances. 1) *Permutation invariance*. Exchanging the indices of any pair of atoms will not change the material. 2) *Translation invariance*. Translating the atom coordinates \mathbf{X} by an arbitrary vector will not change the material. 3) *Rotation invariance*. Rotating the atom coordinates \mathbf{X} and \mathbf{L} together by an arbitrary rotation matrix will not

change the material. 4) *Periodic invariance*. There are infinite different ways of choosing unit cells with different shapes and sizes, for example obtaining a bigger unit cell as an integer multiplier of a smaller unit cell using integer translations. The material will again not change given different choice of unit cells.

D Detailed Methods

D.1 Periodic material encoder.

$\text{PGNN}_{\text{ENC}}(\mathcal{M})$ encodes a material \mathcal{M} as a latent representation $\mathbf{z} \in \mathbb{R}^D$ following the reparameterization trick in VAE [5]. We use a SE(3) invariant periodic graph neural network to encode the material using its multi-graph representation (refer to Appendix B).

D.2 Prediction of aggregated properties.

$\text{MLP}_{\text{AGG}}(\mathbf{z})$ predicts 3 aggregated properties of the encoded material from its latent representation \mathbf{z} . It is parameterized by 3 separate multilayer perceptrons (MLPs). 1) Composition $\mathbf{c} \in \mathbb{R}^{|\mathcal{A}|}$ is predicted by minimizing the cross entropy between the ground truth composition and predicted composition, i.e. $-\sum_i \mathbf{p}_i \log \mathbf{c}_i$. 2) Lattice $\mathbf{L} \in \mathbb{R}^{3 \times 3}$ is reduced to 6 unique, rotation invariant parameters with the Niggli algorithm [23], i.e., the lengths of the 3 lattice vectors, the angles between them, and the values are predicted with an MLP after being normalized to the same scale (Appendix F.1) with an L_2 loss. 3) Number of atoms $N \in \{1, 2, \dots\}$ is predicted with a softmax classification loss from the set of possible number of atoms. \mathcal{L}_{AGG} is a weighted sum of the above 3 losses.

D.3 Conditional score matching decoder.

$\text{PGNN}_{\text{DEC}}(\tilde{\mathcal{M}}|\mathbf{z})$ is a PGNN that inputs a noisy material $\tilde{\mathcal{M}}$ with type noises $\sigma_{\mathcal{A}}$, coordinate noises $\sigma_{\mathcal{X}}$, as well as a latent \mathbf{z} , and outputs 1) a score $\mathbf{s}_{\mathcal{X}}(\tilde{\mathcal{M}}|\mathbf{z}; \sigma_{\mathcal{A}}, \sigma_{\mathcal{X}}) \in \mathbb{R}^{N \times 3}$ to denoise the coordinate for each atom towards its ground truth value, and 2) a probability distribution of the true atom types $\mathbf{p}_{\mathcal{A}}(\tilde{\mathcal{M}}|\mathbf{z}; \sigma_{\mathcal{A}}, \sigma_{\mathcal{X}}) \in \mathbb{R}^{N \times |\mathcal{A}|}$. We use a SE(3) graph network to ensure the equivariance of $\mathbf{s}_{\mathcal{X}}$ with respect to the rotation of $\tilde{\mathcal{M}}$. To obtain the noisy structures $\tilde{\mathcal{M}}$, we sample $\sigma_{\mathcal{A}}$ and $\sigma_{\mathcal{X}}$ from two geometric sequences of the same length: $\{\sigma_{\mathcal{A},j}\}_{j=1}^L, \{\sigma_{\mathcal{X},j}\}_{j=1}^L$, and add the noises with the following methods. For type noises, we use the type distribution defined by the predicted composition \mathbf{c} to linearly perturb true type distribution $\tilde{\mathbf{A}} \sim (\frac{1}{1+\sigma_{\mathcal{A}}} \mathbf{p}_{\mathcal{A}} + \frac{\sigma_{\mathcal{A}}}{1+\sigma_{\mathcal{A}}} \mathbf{p}_{\mathbf{c}})$, where $\mathbf{p}_{\mathcal{A},ij} = 1$ if atom i has the true atom type j and $\mathbf{p}_{\mathcal{A},ij} = 0$ for all other j s, and $\mathbf{p}_{\mathbf{c}}$ is the predicted composition. For coordinate noises, we add Gaussian noises to the true coordinates $\tilde{\mathbf{X}} \sim \mathcal{N}(\mathbf{X}, \sigma_{\mathcal{X}}^2 \mathbf{I})$.

$\text{PGNN}_{\text{DEC}}(\tilde{\mathcal{M}}|\mathbf{z})$ is parameterized by a SE(3) PGNN that inputs a multi-graph representation (??) of the noisy material structure and the latent representation. The node embedding for node i is obtained by the concatenation of the element embedding of \tilde{a}_i and the latent representation \mathbf{z} , followed by a MLP, $\mathbf{h}_i^0 = \text{MLP}(e_{\mathcal{A}}(\tilde{a}_i) \parallel \mathbf{z})$, where \parallel denotes concatenation of two vectors and $e_{\mathcal{A}}$ is a learned embedding for elements. After K message-passing layers, the PGNN outputs a vector per node that is equivariant to the rotation of $\tilde{\mathcal{M}}$. These vectors are used to predict the scores, and we follow [6, 24] to parameterize the score network with noise scaling: $\mathbf{s}_{\mathcal{X}}(\tilde{\mathcal{M}}|\mathbf{z}; \sigma_{\mathcal{A}}, \sigma_{\mathcal{X}}) = \mathbf{s}_{\mathcal{X}}(\tilde{\mathcal{M}}|\mathbf{z})/\sigma_{\mathcal{X}}$. The node representations \mathbf{h}_i^K are used to predict the distribution of true atom types, and the type predictor is the same at all noise levels: $\mathbf{p}_{\mathcal{A}}(\tilde{\mathcal{M}}|\mathbf{z}; \sigma_{\mathcal{A}}, \sigma_{\mathcal{X}}) = \mathbf{p}_{\mathcal{A}}(\tilde{\mathcal{M}}|\mathbf{z}), \mathbf{p}_{\mathcal{A}}(\tilde{\mathcal{M}}|\mathbf{z})_i = \text{softmax}(\text{MLP}(\mathbf{h}_i^K))$.

D.4 Periodicity influences denoising target.

Due to periodicity, a specific atom i may move out of the unit cell defined by \mathbf{L} when the noise is sufficiently large. This leads to two different ways to define the scores for node i . 1) Ignore periodicity and define the target score as $\mathbf{x}_i - \tilde{\mathbf{x}}_i$; or 2) Define the target score as the shortest possible displacement between \mathbf{x}_i and $\tilde{\mathbf{x}}_i$ considering periodicity, i.e. $\mathbf{d}_{\min}(\mathbf{x}_i, \tilde{\mathbf{x}}_i) = \min_{k_1, k_2, k_3} (\mathbf{x}_i - \tilde{\mathbf{x}}_i + k_1 \mathbf{l}_1 + k_2 \mathbf{l}_2 + k_3 \mathbf{l}_3)$. We choose 2) because the scores are the same given two different $\tilde{\mathbf{X}}$ that are periodically equivalent, which is mathematically grounded for periodic structures, and empirically results in much more stable training.

The training loss for the decoder \mathcal{L}_{DEC} can be written as,

$$\frac{1}{2L} \sum_{j=1}^L \left[\mathbb{E}_{q_{\text{data}}(M)} \mathbb{E}_{q_{\sigma_{\mathbf{A},j}, \sigma_{\mathbf{X},j}}(\tilde{M}|M)} \left(\left\| \mathbf{s}_{\mathbf{X}}(\tilde{M}|\mathbf{z}) - \frac{\mathbf{d}_{\min}(\mathbf{X}, \tilde{\mathbf{X}})}{\sigma_{\mathbf{X},j}} \right\|_2^2 + \frac{\lambda_a}{\sigma_{\mathbf{A},j}} \mathcal{L}_a(\mathbf{p}_{\mathbf{A}}(\tilde{M}|\mathbf{z}), \mathbf{p}_{\mathbf{A}}) \right) \right], \quad (2)$$

where λ_a denotes a coefficient for balancing the coordinate and type losses, \mathcal{L}_a denotes the cross entropy loss over atom types, $\mathbf{p}_{\mathbf{A}}$ denotes the true atom type distribution. Note that to simplify the equation, we follow the loss coefficients in [6] for different $\sigma_{\mathbf{X},j}$ and $\sigma_{\mathbf{A},j}$ and factor them into Equation 2.

Algorithm 1 Material Generation via Annealed Langevin Dynamics

- 1: **Input:** latent representation \mathbf{z} , type and coordinate noise levels $\{\sigma_{\mathbf{A}}\}$, $\{\sigma_{\mathbf{X}}\}$, step size ϵ , number of sampling steps T
 - 2: Predict aggregated properties \mathbf{c} , \mathbf{L} , N from \mathbf{z} .
 - 3: Uniformly initialize \mathbf{X}_0 within the unit cell by \mathbf{L} .
 - 4: Randomly initialize \mathbf{A}_0 with \mathbf{c} .
 - 5: **for** $j \leftarrow 1$ to L **do**
 - 6: $\alpha_j \leftarrow \epsilon \cdot \sigma_{\mathbf{X},j}^2 / \sigma_{\mathbf{X},L}^2$
 - 7: **for** $t \leftarrow 1$ to T **do**
 - 8: $\mathbf{s}_{\mathbf{X},t} \leftarrow \mathbf{s}_{\mathbf{X}}(\mathbf{A}_{t-1}, \mathbf{X}_{t-1}, \mathbf{L}|\mathbf{z}; \sigma_{\mathbf{A},j}, \sigma_{\mathbf{X},j})$
 - 9: $\mathbf{p}_{\mathbf{A},t} \leftarrow \mathbf{p}_{\mathbf{A}}(\mathbf{A}_{t-1}, \mathbf{X}_{t-1}, \mathbf{L}|\mathbf{z}; \sigma_{\mathbf{A},j}, \sigma_{\mathbf{X},j})$
 - 10: Draw $\mathbf{X}_t^\epsilon \sim \mathcal{N}(0, \mathbf{I})$
 - 11: $\mathbf{X}'_t \leftarrow \mathbf{X}_{t-1} + \alpha_j \mathbf{s}_{\mathbf{X},t} + \sqrt{2\alpha_j} \mathbf{X}_t^\epsilon$
 - 12: $\mathbf{X}_t \leftarrow \text{back_to_cell}(\mathbf{X}'_t, \mathbf{L})$
 - 13: $\mathbf{A}_t = \text{argmax} \mathbf{p}_{\mathbf{A},t}$
 - 14: $\mathbf{X}_0 \leftarrow \mathbf{X}_T, \mathbf{A}_0 \leftarrow \mathbf{A}_T$
-

D.5 Material generation with Langevin dynamics.

After training the model, we can generate the periodic structure of material given a latent representation \mathbf{z} . First, we use \mathbf{z} to predict the aggregated properties: 1) composition \mathbf{c} , 2) lattice \mathbf{L} , and 3) the number of atoms N . Then, we randomly initialize an initial periodic structure $(\mathbf{A}_0, \mathbf{X}_0, \mathbf{L})$ with the aggregated properties and perform an annealed Langevin dynamics [6] using the decoder, simultaneously updating the atom types and coordinates. During the coordinate update, we map the coordinates back to the unit cell at each step if atoms move out of the cell. The algorithm is summarized in Algorithm 1.

E Connection between the gradient field and a harmonic force field.

The gradient field $\mathbf{s}_{\mathbf{X}}(\tilde{M}|\mathbf{z})$ is used to update atom coordinates in Langevin dynamics via the force term, $\alpha_j \mathbf{s}_{\mathbf{X},t}$. Below, we show that $\alpha_j \mathbf{s}_{\mathbf{X},t}$ is mathematically equivalent to⁶ a harmonic force field $F(\tilde{\mathbf{X}}) = -k(\tilde{\mathbf{X}} - \mathbf{X})$ when the noises are small, where \mathbf{X} is the equilibrium position of the atoms and k is a force constant. Harmonic force field, i.e. spring-like force field, is a simple yet general physical model that approximates the forces on atoms when they are close to their equilibrium locations. This indicates that our learned gradient field utilizes the harmonic approximation to approximate QM forces without any explicit force data and generates stable materials with this physically motivated inductive bias.

To prove the above statement, we assume the loss in Equation 2 can be minimized to zero when the noises are small, meaning that

$$\mathbf{s}_{\mathbf{X}}(\tilde{\mathbf{A}}, \tilde{\mathbf{X}}, \mathbf{L}|\mathbf{z}) = \frac{\mathbf{d}_{\min}(\mathbf{X}, \tilde{\mathbf{X}})}{\sigma_{\mathbf{X},j}}, \forall j > J, \quad (3)$$

where $\sigma_{\mathbf{X},j} \in \{\sigma_{\mathbf{X},j}\}_{j=1}^L$ and any noise smaller than $\sigma_{\mathbf{X},j}$ is considered as small.

⁶In fact, this is also true for the original formulation of NCSN [6]

The force term in the Langevin dynamics $\alpha_j \mathbf{s}_{\mathbf{X},t}$ can then be written as

$$\alpha_j \mathbf{s}_{\mathbf{X}}(\tilde{\mathbf{A}}, \tilde{\mathbf{X}}, \mathbf{L} | \mathbf{z}; \sigma_{\mathbf{A},j}, \sigma_{\mathbf{X},j}) = \epsilon \cdot \sigma_{\tilde{\mathbf{X}},j}^2 / \sigma_{\tilde{\mathbf{X}},L}^2 \cdot \mathbf{s}_{\mathbf{X}}(\tilde{\mathbf{A}}, \tilde{\mathbf{X}}, \mathbf{L} | \mathbf{z}) / \sigma_{\mathbf{X},j} \quad (4)$$

$$= \epsilon \cdot \frac{\sigma_{\tilde{\mathbf{X}},j}^2}{\sigma_{\tilde{\mathbf{X}},L}^2} \cdot \frac{\mathbf{d}_{\min}(\mathbf{X}, \tilde{\mathbf{X}})}{\sigma_{\tilde{\mathbf{X}},j}^2}, \forall j > J \quad (5)$$

$$= -\frac{\epsilon}{\sigma_{\tilde{\mathbf{X}},L}^2} \mathbf{d}_{\min}(\tilde{\mathbf{X}}, \mathbf{X}), \forall j > J \quad (6)$$

If we write $\epsilon / \sigma_{\tilde{\mathbf{X}},L}^2 = k$, then,

$$\alpha_j \mathbf{s}_{\mathbf{X}}(\tilde{\mathbf{A}}, \tilde{\mathbf{X}}, \mathbf{L} | \mathbf{z}; \sigma_{\mathbf{A},j}, \sigma_{\mathbf{X},j}) = -k \mathbf{d}_{\min}(\tilde{\mathbf{X}}, \mathbf{X}), \forall j > J \quad (7)$$

If the noises are small enough that atoms do not cross the periodic boundaries, then we have $\mathbf{d}_{\min}(\mathbf{X}, \tilde{\mathbf{X}}) = \mathbf{X} - \tilde{\mathbf{X}}$. Therefore,

$$\alpha_j \mathbf{s}_{\mathbf{X}}(\tilde{\mathbf{A}}, \tilde{\mathbf{X}}, \mathbf{L} | \mathbf{z}; \sigma_{\mathbf{A},j}, \sigma_{\mathbf{X},j}) = -k(\tilde{\mathbf{X}} - \mathbf{X}), \forall j > J. \quad (8)$$

F Implementation details

F.1 Prediction of lattice parameters

There are infinitely many different ways of choosing the lattice for the same material. We compute the Niggli reduced lattice [23] with pymatgen [25], which is a unique lattice for any given material. Since the lattice matrix \mathbf{L} is not rotation invariant, we instead predict the 6 lattice parameters, i.e. the lengths of the 3 lattice vectors and the angles between them. We normalize the lengths of lattice vectors with $\sqrt[3]{N}$, where N is the number of atoms, to ensure that the lengths for materials of different sizes are at the same scale.

F.2 Multi-graph construction

For the encoder, we use CrystalNN [22] to determine edges between atoms and build a multi-graph representation. For the decoder, since it inputs a noisy structure generated on the fly, the multi-graph must also be built on the fly for both training and generation, and CrystalNN is too slow for that purpose. We use a KNN algorithm that considers periodicity to build the decoder graph where $K = 20$ in all of our experiments.

F.3 GNN architecture

We use DimeNet++ adapted for periodicity [26, 27] as the encoder, which is SE(3) invariant to the input structure. The decoder needs to output a vector per node that is SE(3) equivariant to the input structure. We use GemNet-dQ [28] as the decoder. We used implementations from the Open Catalysis Project (OCP) [29], but we reduced the size of hidden dimensions to 128 for faster training. The encoder has 2.2 million parameters and the decoder has 2.3 million parameters.

G Evaluation metrics

G.1 Reconstruction

We use StructureMatcher from pymatgen [25], which finds the best match between two structures considering all invariances of materials. The match rate is the percentage of materials satisfying the criteria `stol=0.5`, `angle_tol=10`, `ltol=0.3`. The RMSE is averaged over all matched materials. Because the inter-atomic distances can vary significantly for different materials, the RMSE is normalized by $\sqrt[3]{V/N}$, roughly the average atom radius per material.

G.2 Generation

Validity. Following [7], a structure is valid as long as the shortest distance between any pair of atoms is larger than 0.5 Å, which is a relative weak criterion. The composition is valid if the overall charge is neutral as computed by SMOG [30].

Coverage (COV). Inspired by [31, 32], we define two coverage metrics, COV-R (Recall) and COV-P (Precision), to measure the similarity between ensembles of generated materials and ground truth materials in test set. Intuitively, COV-R measures the percentage of ground truth materials being correctly predicted, and COV-P measures the percentage of predicted materials having high quality.

Property statistics. We compute the earth mover’s distance (EMD) between the property distribution of generated materials and test materials. We use density (ρ , unit g/cm^3), energy predicted by an independent GNN (E , unit eV/atom), and number of unique elements (# elem.) as our properties. Validity and coverage are computed over 10,000 materials randomly sampled from $\mathcal{N}(0, 1)$. Property statistics is computed over 1,000 valid materials randomly sampled from those that pass the validity test.

G.3 Property Optimization

Following [33], we jointly train a property predictor F parameterized by an MLP to predict properties of training materials from latent z . To optimize properties, we start with the latent representations of testing materials and apply gradient ascent in the latent space to improve the predicted property $F(\cdot)$. After applying 5000 gradient steps with step sizes of 1×10^{-3} , 10 materials are decoded from the latent trajectories every 500 steps. We use an independently trained property predictor to select the best one from the 10 decoded materials. Cond-DFC-VAE is a conditional VAE so we directly condition on the target property, sample 10 materials, and select the best one using the property predictor.

H Supplementary Figures

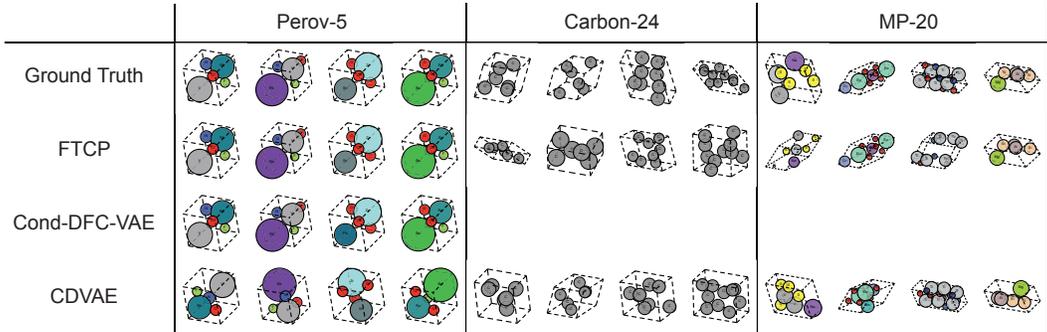


Figure 3: Reconstructed structures of randomly selected materials in the test set.

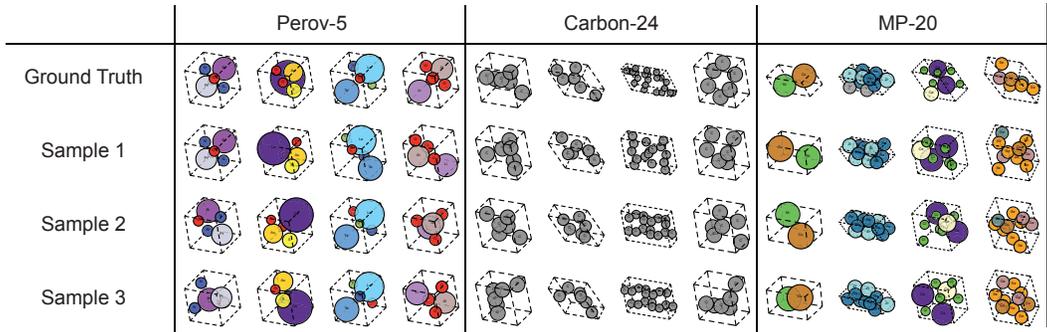


Figure 4: Different reconstructed structures from CDVAE from the same z , following 3 Langevin dynamics sampling with different random seeds.

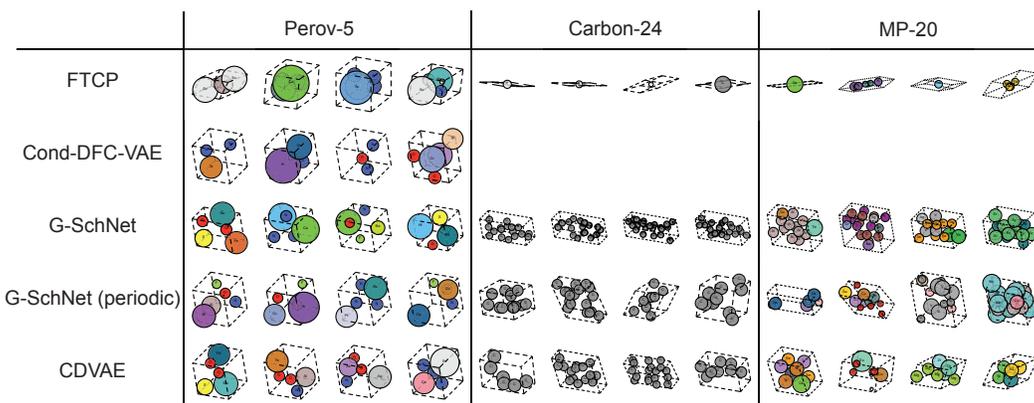


Figure 5: Structures sampled from $\mathcal{N}(0, 1)$ for the generation task.