WyckoffTransformer: Generation of Symmetric Crystals

Nikita Kazeev[a], A. Ustyuzhanin[ba], I. Romanov[c], R. Zhu[d], W. Nong[d], S. Yamazaki[d], K. Hippalgaonkar[d]

[a] Institute for Functional Intelligent Materials, National University of Singapore, 4 Science Drive 2, Singapore 117544
[b] Constructor University, Campus Ring 1, Bremen 28759, Germany
[c] Tsinghua University Haidian District, 100084, Beijing, Chin
[d] School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798

TLDR

A generative model for materials: permutation–invariant autoregressive generative Transformer encoder with a symmetry-based representation

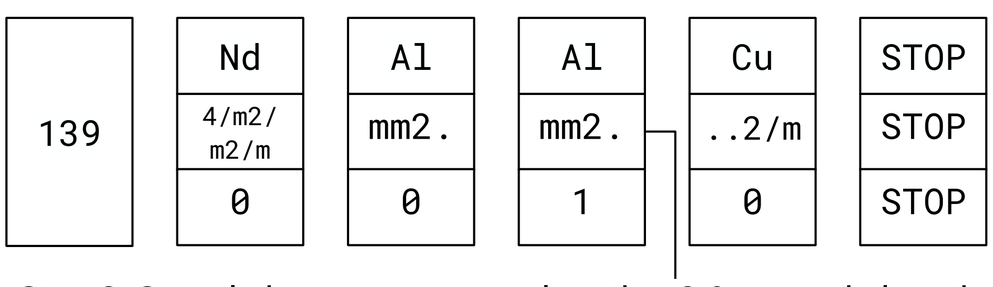
Representation

WyFormer samples coordinate–free Wyckoff representations conditioned on space group symmetry. Full structures can then be constructed with DiffCSP++ [1] or pyXtal [8] & MLIP.

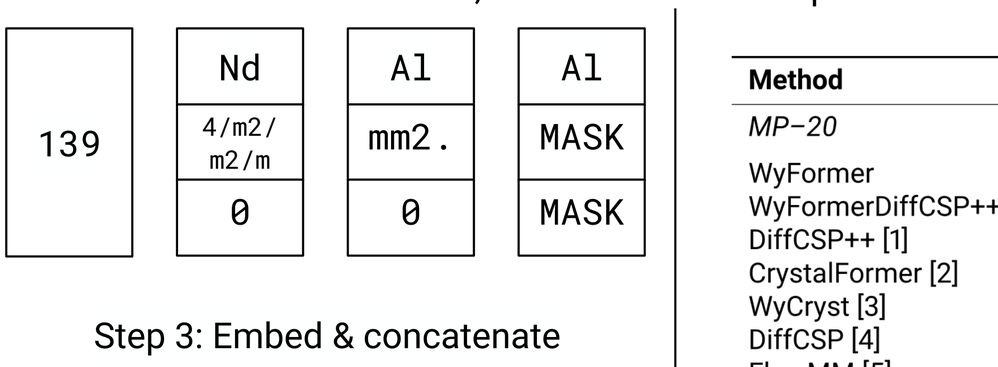
```
Composition: Nd2Al16Cu8
Group: I 4/m m m (139)
  8.9013,   8.9013,   5.1991,  90.0000,  90.0000,  90.0000
Wyckoff sites:
    Nd @ [ 0.0000  0.0000  0.0000], WP [2a] Site [4/m2/m2/m]
    Al @ [ 0.2788  0.5000  0.0000], WP [8j] Site [mm2.]
    Al @ [ 0.6511  0.0000  0.0000], WP [8i] Site [mm2.]
    Cu @ [ 0.2500  0.2500  0.2500], WP [8f] Site [..2/m]
```

Model & Training

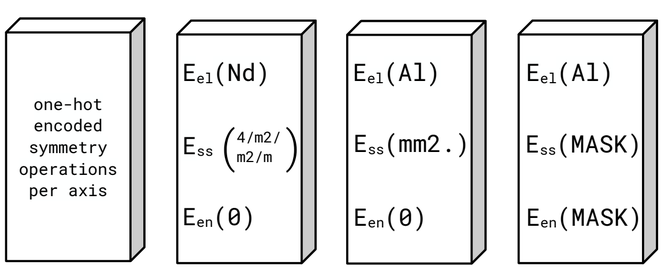
Step 1: Tokenize



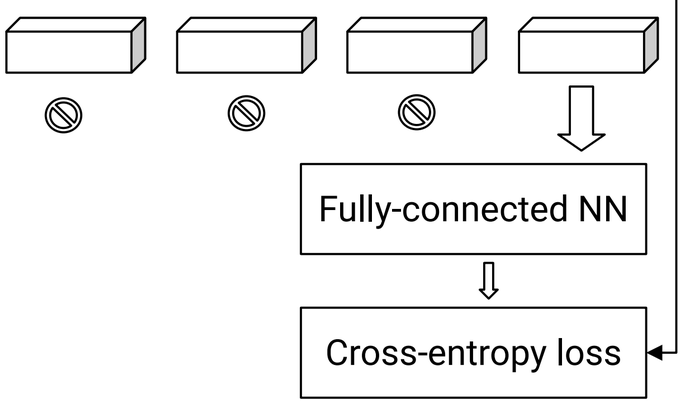
Step 2: Sample known sequence length = 3 & cascade length = 1
Remove the unknown tokens, MASK the unknown parts of tokens



Step 3: Embed & concatenate

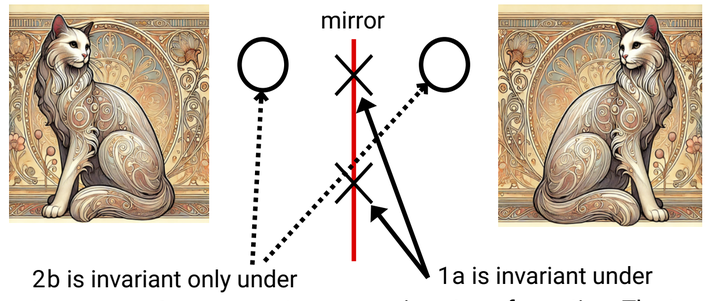


Step 4: Transformer encoder

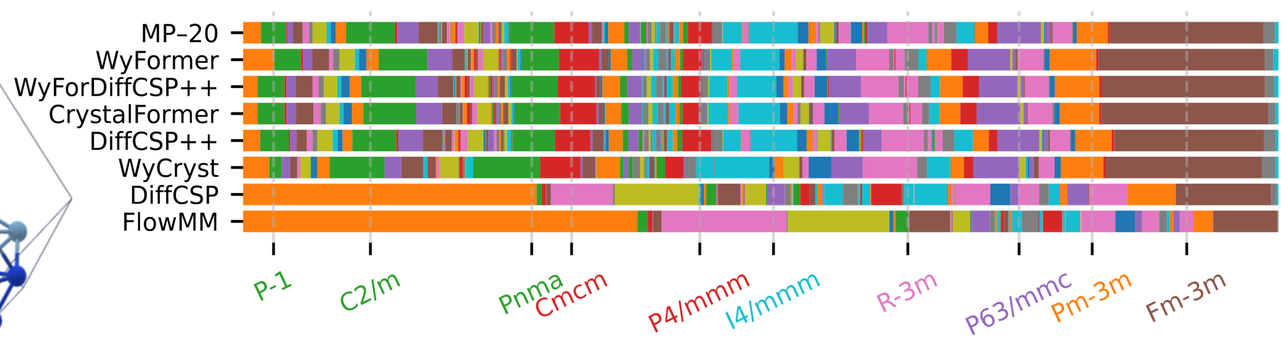


Wyckoff Positions

For a given symmetry transformation some points are special: they remain invariant. For a given space group, a set of points invariant under given transformation is called a Wyckoff position



Nature is Symmetric



Distribution of space groups in MP–20 dataset and the generated samples. 98% of MP–20 [6] structures belong to symmetry groups other than P1, which consists of only lattice transition. For symmetry-unaware models, DiffCSP & FlowMM, just 63% and 56% do → their structures are unrealistic!

Generation Evaluation

Method	Novel Unique	P1 (%)	Space Group	DFT ↑			r	CHGNet ↑	
MP–20	Templates (#) ↑		χ^2 ↓	#	S.U.N. (%)	S.S.U.N. (%)		S.U.N. (%)	S.S.U.N. (%)
WyFormer	180	3.24	0.223	96	7.5	7.5	0.33	39.2	38.2
WyFormerDiffCSP++	**186**	**1.46**	**0.212**	95	**14.1**	**14.1**	0.44	36.7	36.0
DiffCSP++ [1]	*10*	2.57	0.255	94	8.5	**8.5**	0.32	41.4	**40.8**
CrystalFormer [2]	74	0.91	0.276	–	–	–	–	33.9	33.8
WyCryst [3]	165	4.79	0.710	–	–	–	–	36.6	35.2
DiffCSP [4]	76	36.57	7.989	82	**20.8**	**13.1**	0.36	**57.4**	40.6
FlowMM [5]	51	44.27	12.423	–	–	–	–	49.2	29.9
WyFormer *MPTS–52*	386	0	0.225	–	–	–	–	24.4	24.4

S.U.N. = Stable Unique Novel; S.S.U.N. = Symmetric Stable Unique Novel

Property Prediction

Method	Energy, meV	Band gap, meV	Train	Test
CGCNN	31	292	Materials Project–2018.6.1	
SchNet	33	345		
MEGNet	30	307		
GATGNN	33	280		
ALIGNN	22	218		
Matformer	21	211		
PotNet	**19**	**204**		
CHGNet	34	–	MPTrj	MP–20
WyFormer	25	234	MP–20	

AFLOW

Method	Thermal conductivity	Debye temperature	Bulk modulus	Shear modulus
Roost	2.70	37.17	8.82	9.98
CrabNet	2.32	**33.46**	**8.69**	**9.08**
HotCrab	2.25	35.76	9.10	9.43
ElemNet	3.32	45.72	12.12	13.32
RF	2.66	36.48	11.91	10.09
WyFormer	**2.20**	36.36	9.63	10.14

Advertisements

Come to Singapore! It is the most expensive city in the world – because it is worth it!

Attend AI4X (not *that* 𝕏) conference 8 to 11 July 2025. Covers AI for most natural sciences and, eclectically, finance. Abstract submission deadline ≈February 1.

Attend the ICLR 2025 workshop on multi scale ML on April 27/28 2025. Given low-level theory and code, how can we model complex systems on a useful time scale?

Apply for postdoc & PhD student positions at the multi scale ML project at the National University of Singapore

References

[1] Jiao, Rui, et al. "Space Group Constrained Crystal Generation." The Twelfth International Conference on Learning Representations
[2] Cao, Zhendong, et al. "Space Group Informed Transformer for Crystalline Materials Generation." arXiv preprint arXiv:2403.15734 (2024).
[3] Zhu, Ruiming, et al. "WyCryst: Wyckoff inorganic crystal generator framework." Available at SSRN 4658842 (2023).
[4] Jiao, Rui, et al. "Crystal structure prediction by joint equivariant diffusion." NeurIPS 2024
[5] Miller, Benjamin Kurt, et al. "FlowMM: Generating Materials with Riemannian Flow Matching." ICML 2024
[6] Xie, Tian, et al. "Crystal Diffusion Variational Autoencoder for Periodic Material Generation." ICLR 2022
[7] Curtarolo, Stefano, et al. "AFLOW: An automatic framework for high-throughput materials discovery." Computational Materials Science 58 (2012): 218-226.
[8] Fredericks, Scott, et al. "PyXtal: A Python library for crystal structure generation and symmetry analysis." Computer Physics Communications 261 (2021): 107810.

Conclusion

WyFormer has a unique and powerful synergy of attributes, proven by extensive experimentation: best–in–class symmetry–conditioned generation, physics–motivated inductive bias, competitive stability of the generated structures, competitive property prediction quality, and unparalleled inference speed. Altogether, they make it a great starting point in the material design pipeline.

Paper Code Contact