A generative material transformer using Wyckoff representation

Pierre-Paul De Breuck[1], Hashim A. Piracha[1], Gian-Marco Rignanese[2], Miguel A. L. Marques[1]
[1]RC FEMS, ICAMS, RUB, Bochum, Germany
[2]UCLouvain, Louvain-la-Neuve, Belgium

pierrre-paul.debreuck@uclouvain.be

Identifying and enumerating novel stable compounds, those near the convex hull, is essential for many technological applications. In this I will present present Matra-Genoa, an autoregressive transformer model built on invertible tokenized representations of symmetrized crystals, including free coordinates. This approach enables sampling from a hybrid action space. The model is trained across the periodic table and space groups and can be conditioned on specific properties. We demonstrate its ability to generate stable, novel, and unique crystal structures by conditioning on the distance to the convex hull. Resulting structures are 8 times more likely to be stable than baselines using PyXtal with charge compensation, while maintaining high computational efficiency. I'll also explain how Matra-Genoa contributes to the Alexandria pipeline for the discovery of novel compounds.

References

[1] De Breuck, PP., Piracha, H.A., Rignanese, GM. et al. A generative material transformer using Wyckoff representation. npj Comput Mater 12, 60 (2026).

Figures

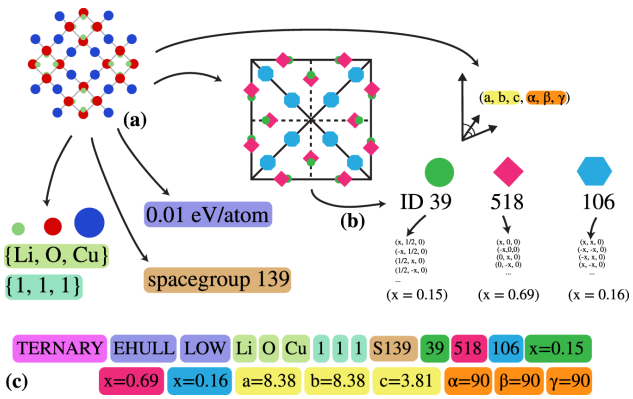

Figure 1. Schematic overview of the invertible sequenced representation