

Master thesis proposal

Benchmark of generative AI models for crystal structure prediction

Keywords : Crystal Structure Prediction (CSP), Generative AI, Generative Adversarial Network, Variational Auto-Encoder, Diffusion Model

SCIENTIFIC DESCRIPTION :

Crystal structure prediction (CSP) consists in determining the crystal structure of all the (stable and metastable) polymorphs of an inorganic compound for which only the chemical formula is provided [1]. Widely applied upstream of the exploratory synthesis of new compounds (computational design of new materials), it can also be used to assist experimental characterization of newly synthesized compounds where crystal structure resolution may prove difficult, either because the sample is metastable, poorly crystallized, decomposes under the beam, or presents contradictory structural characteristics depending on the characterization method.

For almost 20 years, these problems have been solved using global optimizers of the compound energy landscape (genetic/evolutionary algorithms, particle swarms, random search methods) and have generally been successful, particularly in predicting structures unknown to current databases and original according to accepted knowledge [2,3]. For 3 years now, algorithms based on Generative Artificial Intelligence (Gen AI) have sought to compete with these approaches, highlighting numerous advantages (computation time, diversity of structures, transferability, ...) but no benchmark dedicated to materials science currently allows a critical analysis of these approaches. We have recently shown, for example, that crystallochemical constraints can penalize generative antagonistic networks [4]. The aim of this M2 project is to compare the various Gen AI models proposed for crystal structure prediction in materials science, and to provide a critical analysis using benchmarks. This will involve

- 1) on the one hand, to take stock of the databases useful for testing, and to propose and build additional ones if necessary,
- 2) select the criteria for characterizing these methods (chemical validity, error rate, consideration of energy classification/relaxation requirements, calculation time, etc.).
- 3) test these different algorithms on the same basis.

[1] *Modern Methods of Crystal Structure Prediction*, A.R. Oganov (editor), Wiley (2010)

[2] S. Sasaki, M. T. Caldes, C. Guillot-Deudon, I. Braems et al, Design of metastable oxychalcogenide phases by topochemical (de)intercalation of sulfur in $\text{La}_2\text{O}_2\text{S}_2$, *Nat. Comm.* 12, 3605 (2021)

[3] B. Wang, I. Braems, S. Sasaki et al, Prediction of a New Layered Polymorph of FeS_2 with $\text{Fe}^{3+}\text{S}^{2-}(\text{S}_2^{2-})_{1/2}$ Structure, *The Journal of Physical Chemistry Letters* 11(20), (2020)

[4] G. Pertus, Etude et mise en place de méthodes génératives pour la prédiction de structures cristallines, rapport de stage, Nantes Université, (2024)

Techniques/methods in use: The student recruited will have access to the GENCI supercomputer dedicated to AI (Jean Zay) for the training and generation phases (on GPU), and to the IMN local cluster and the GliCID mesoscale computer. All sources are written in Python. Depending on the investigated code, structure energy may not be embedded in the structure representation. For these cases, the energy of generated structures will be computed via a DFT procedure/workflow using VASP.



Master thesis proposal

Budget description and use: The grant will mostly finance the internship bonus for a Master 2 student and potential environment requirements (generative AI and architecture optimization formation (GENCI) and workshop (GDR IAMAT))

Applicant skills: Strong Python skills and motivation are essential. A solid grounding in the basics of AI (at least discriminative: machine/deep learning) is recommended. Knowledge of materials science or crystallochemistry is a very welcome bonus, but not mandatory.

Industrial partnership: This is a pioneer project on Generative AI for which no industrial partner is formally involved (yet) as the supervisor is convinced that these methods need this kind of work to emerge efficiently. Nevertheless the internship supervisor already has several industrial collaborations among which she holds a TANDEM PhD these on CSPs and we hope the resulted benchmark of this M2 project will motivate collaborations on this emerging field.

Internship supervisor(s): Isabelle BRAEMS-ABBASPOUR, 02 40 33 69 33, Isabelle.braems@cnrs-imn.fr

Internship location: Institut des Matériaux de Nantes Jean Rouxel, 2 rue de la Houssinière, Nantes, France.

Possibility for a Doctoral thesis: Not Yet defined.

