Generative Modeling for Materials Design and Discovery using Invertible String-based Materials Representation

Viejay Ordillo
Nara Institute of Science and Technology

Inverse materials design has traditionally relied on expert intuition and high-throughput virtual screening. Recent advancements in machine learning have brought generative deep learning models to the forefront, particularly those that represent crystal structures as graphs. However, directly processing 3D atomic graphs or point clouds often incurs substantial computational and storage overhead. Alternatively, string-based representations offer a compact and efficient format, though their application to extended systems such as crystalline solids remain limited. In molecular science, line notations like SMILES and SELFIES have enabled powerful chemical language models, whereas materials science has only recently introduced the Simplified Line-Input Crystal-Encoding System (SLICES) as an analogous format for crystalline materials. To date, SLICES has been employed solely in MatterGPT, a transformer-based model, and its broader utility in generative modeling for materials remains largely unexplored. In this study, we systematically investigate the use of SLICES as a unified string-based representation for materials design. We benchmark several generative deep learning architectures, assessing their effectiveness in generating chemically valid and diverse materials. We also aim to provide a foundation for using SLICES in materials discovery and to identify the main challenges that need to be addressed to advance its applications.