Towards invertible 2D crystal structure representation for efficient downstream task execution

Andrey Ustyuzhanin[1,2], Egor Shibaev[1]
[1]Constructor University 28759 Bremen, Germany
[2]Institute for Functional Intelligent Materials, National University of Singapore, 117544, Singapore

andrey.ustyuzhanin@constructor.org

The field of material science has undergone a profound transformation with the emergence of computational methods, enabling the exploration and design of novel materials with customized properties. Among these materials, two-dimensional (2D) crystals, particularly transition metal dichalcogenides (TMDCs) like molybdenum disulfide MoS2, have received significant attention due to their unique electronic, optical, and mechanical properties Tedstone et al. [2015], Wang et al. [2012], Han et al. [2015]. These properties not only make them suitable for a wide range of applications, from transistors to photovoltaic devices but also provide a fertile ground for fundamental research, specifically in comprehending and manipulating crystal defects. Recently, controllable defect engineering and formation of alloys became possible, as shown in Huang et al. [2023]. These methods are now viable tools to prepare new 2D materials with predetermined properties. The monolayer of MoS2, a prominent member of the two-dimensional materials family, is composed of a plane of molybdenum atoms flanked by sulfur atom planes. However, accurately analyzing defect arrangements computationally can be challenging, particularly considering crystalline symmetries. As a result, there is an urgent need for computational tools that can efficiently and precisely model, characterize, and predict the effects of defects on material properties.

In the study of MoS2 lattice defects, we explore the use of Siamese Neural Networks (SNN) [Bromley et al. [1993]] to create invariant embeddings, which respect the crystalline symmetry of the lattice. By training our model with contrastive learning, we successfully differentiate configurations with varying defects, achieving perfect accuracy in recognizing equivalent placements. Our method showcases the capability to predict physical properties like formation energy per site and the bandgap with strong performance across both low and high-defect density scenarios, outperforming traditional methods when enhanced with polynomial features. Despite its effectiveness, the model presents limitations at high defect densities, indicating a need for further refinement. Our approach lays the groundwork for reverse-engineering processes. Thus, we open pathways for generative models that can navigate from specified property ranges to optimal defect configurations, fostering an efficient solution-space exploration for bespoke material synthesis.

References

- A. Tedstone, D. J. Lewis, R. Hao, S.-M. Mao, P. Bellon, R. S. Averback, C. P. Warrens, K. R. West, P. Howard, S. Gaemers, S. J. Dillon, and P. O'Brien. Mechanical properties of molybdenum disulfide and the effect of doping: An in situ tem study. ACS Applied Materials & Interfaces, 7(37):20829–20834, Sep 2015. ISSN 1944-8244. doi: 10.1021/acsami.5b06055.
- Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano. Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. Nature Nanotechnology, 7(11):699–712, 2012. ISSN 1748-3395. doi: 10.1038/nnano.2012.193.
- S. A. Han, R. Bhatia, and S.-W. Kim. Synthesis, properties and potential applications of two-dimensional transition metal dichalcogenides. Nano Convergence, 2(1):17, Sep 2015. ISSN 2196-5404. doi: 10.1186/ s40580-015-0048-4.
- J. Bromley, J. Bentz, L. Bottou, I. Guyon, Y. Lecun, C. Moore, E. Sackinger, and R. Shah. Signature verification using a "siamese" time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence, 7:25, 08 1993.
- P. Huang, R. Lukin, M. Faleev, and et al. Unveiling the complex structure-property correlation of defects in 2d materials based on high throughput datasets. npj 2D Materials and Applications, 7(6), 2023.

Figures

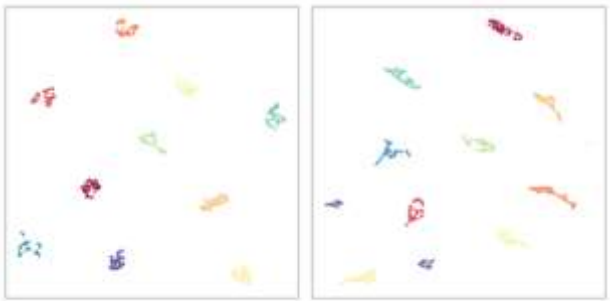

Figure 1. UMAP representations for various defect counts. Each point represents the embedding of some placement. The same color denotes the same configurations.

Method	descriptor length	formation energy per site, MAE, meV		HOMO-LUMO gap, MAE, meV	
		low density	high density	low density	high density
Bag of Bonds	18336	956.8	482.5	612.1	240.2
Coulomb Matrix	36864	184.7	198.8	106.2	159.4
SNN embeddings	12	47.1	382.7	141.4	158.4
SNN embeddings + polynomial features	91	**39.9**	**139.6**	**63.2**	**110.1**

Figure 2. Comparative Performance of Different Methods in Predicting Formation Energy Per Site and HOMO-LUMO Gap. The table compares the MAE values achieved by Bag of Bonds, Coulomb Matrix, and SNN embeddings (with and without polynomial features) across low and high-density datasets.