Reverse Design of High-Activity High-Stability Acidic OER Catalysts Based on MatterGen

Zhihao Wang[a], **Lei Wang***[b] **Xiaonan Wang***[a]

[a] *Department of Chemical Engineering, Tsinghua University, Beijing 100084, China* wangxiaonan@tsinghua.edu.cn

[b] *Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore* wanglei8@nus.edu.sg

1. Introduction

The reverse design of high-activity and high-stability acidic oxygen evolution reaction (OER) catalysts presents a significant challenge, as these materials must simultaneously satisfy thermodynamic stability, resistance to corrosion under low pH conditions, and strong catalytic performance. Traditional experimental and screening methods often encounter limitations such as a restricted sample space and high trial-and-error costs. In contrast, a generative model such as MatterGen [1] offers an innovative solution by directly generating crystal structures through diffusion-based modeling. With the support of large-scale data, MatterGen can learn the distribution of stable materials and precisely control desired properties through conditional fine-tuning. In this way, it provides a powerful framework for identifying new materials that fall below the convex hull energy threshold, maintain corrosion resistance in strongly acidic media, and exhibit high catalytic activity for OER.

2. Methodology

2.1 Comprehensive Dataset Construction

The workflow begins by preparing and constructing a comprehensive dataset. The dataset comprises two parts. The unlabeled dataset, sourced from public databases such as Materials Project [2], Inorganic Crystal Structure Database (ICSD) [3], and Open Quantum Materials Database (OQMD) [4], provides fundamental crystal structure information for pre-training MatterGen to generate valid structures. The labeled dataset is curated from literature on acidic OER catalysts, containing key performance metrics such as adsorption energies for critical intermediates, convex hull energy, overpotential, and degradation time. This part is used to fine-tune the model, ensuring that generated materials not only remain thermodynamically stable but also meet specific catalytic requirements. By matching every catalyst entry with its corresponding crystal structure, we can first build a unified degradation-stability database for OER catalysts under acidic conditions.

2.2 Baseline Model Pre-Training

Once data preparation is complete, the baseline model is pre-trained on the unlabeled dataset. During this phase, MatterGen learns to generate valid crystal structures through its diffusion process by capturing the intrinsic distribution and diversity of crystal configurations. This unsupervised training step establishes a foundational understanding of material structures, which serves as the groundwork for subsequent property-guided fine-tuning.

2.3 Conditional Fine-Tuning and Generative Process

Building on the pre-trained model, adapter modules are integrated to incorporate additional property constraints relevant to acidic OER catalysts. In this fine-tuning stage, metrics such as overpotential, corrosion resistance, and convex hull energy are employed to steer the model toward generating materials that meet specific performance requirements. Classifier-free [5] guidance further allows dynamic control over the enforcement of each property during generation, ensuring that the resulting structures balance thermodynamic stability with optimized catalytic efficacy.

2.4 Candidate Screening

The next stage involves screening and validating the generated candidates. Rapid computational methods—such as machine learning force fields (MLFF) or simplified density functional theory (DFT) calculations—provide an initial check of thermodynamic and catalytic indicators. Only those candidates demonstrating promise across essential benchmarks (below-threshold hull energies, acceptable corrosion properties, and encouraging OER performances) advance to more detailed DFT analyses.

2.5 Experimental Validation

The final phase involves experimental synthesis and electrochemical testing of the top-performing candidates. By evaluating catalytic performance under laboratory conditions, the practical efficacy of each candidate is validated and real-world data on corrosion resistance and overpotential is collected. This experimental feedback is then incorporated into the model to refine its predictive capabilities, guiding subsequent iterations of catalyst design.

3. Conclusion

In summary, the reverse design strategy outlined here leverages a generative model trained on expansive datasets, followed by property-guided fine-tuning and comprehensive validation. By systematically combining thermodynamic stability criteria, acid-resistant performance indicators, and catalytic efficiency requirements, the method targets materials that not only resist corrosion in highly acidic environments but also drive OER at lower overpotentials. This approach underscores the effectiveness of MatterGen in condition-specific crystal structure generation, offering a new avenue for rational catalyst design. Beyond providing viable candidate catalysts for acid-based OER, this workflow can be generalized to other advanced materials discovery tasks, illustrating the potential of data-driven generative modeling to accelerate innovation in materials science.

Reverse Design of High-Activity High-Stability Acidic OER Catalysts Based on MatterGen

Zhihao Wang[a], **Lei Wang***[b] **Xiaonan Wang***[a]

[a] *Department of Chemical Engineering, Tsinghua University, Beijing 100084, China* wangxiaonan@tsinghua.edu.cn

[b] *Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore* wanglei8@nus.edu.sg

Acknowledgments

This work was supported by the National Key R&D Program of China (2022ZD0117501), and Tsinghua University Initiative Scientific Research Program.

References

[1] C. Zeni et al., A generative model for inorganic materials design. Nature, Jan. 2025, doi: 10.1038/s41586-025-08628-5.

[2] A. Jain et al., Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Materials, vol. 1, no. 1, p. 011002, Jul. 2013, doi: 10.1063/1.4812323.

[3] D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, and S. Rehme, Recent developments in the Inorganic Crystal Structure Database: theoretical crystal structure data and related features. J Appl Crystallogr, vol. 52, no. 5, pp. 918–925, Oct. 2019, doi: 10.1107/S160057671900997X.

[4] S. Kirklin et al., The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. npj Comput Mater, vol. 1, no. 1, p. 15010, Dec. 2015, doi: 10.1038/npjcompumats.2015.10.

[5] J. Ho and T. Salimans, Classifier-Free Diffusion Guidance. Jul. 26, 2022, arXiv: arXiv:2207.12598. doi: 10.48550/arXiv.2207.12598.