Catalyst GFlowNet for electrocatalyst design: A hydrogen evolution reaction case study

Lena Podina[1,2] **Alex Hernández-García**[1,3] **Christina Humer**[4] **Alexandre Duval**[5]
Victor Schmidt[5] **Ali Ramlaoui**[5] **Shahana Chatterjee**[1] **Yoshua Bengio**[1,3]
David Rolnick[1,6] **Félix Therrien**[1]
[1]Mila Quebec AI Institute [2]University of Waterloo [3]Université de Montréal
[4]ETH Zurich [5]Entalpic [6]McGill University
lpodina@uwaterloo.ca
felix.therrien@mila.quebec

Abstract

Efficient and inexpensive energy storage is essential for accelerating the adoption of renewable energy and ensuring a stable supply, despite fluctuations in sources such as wind and solar. Electrocatalysts play a key role in hydrogen energy storage (HES), allowing the energy to be stored as hydrogen. However, the development of affordable and high-performance catalysts for this process remains a significant challenge. We introduce Catalyst GFlowNet, a generative model that leverages machine learning-based predictors of formation and adsorption energy to design crystal surfaces that act as efficient catalysts. We demonstrate the performance of the model through a proof-of-concept application to the hydrogen evolution reaction, a key reaction in HES, for which we successfully identified platinum as the most efficient known catalyst. In future work, we aim to extend this approach to the oxygen evolution reaction, where current optimal catalysts are expensive metal oxides, and open the search space to discover new materials. This generative modeling framework offers a promising pathway for accelerating the search for novel and efficient catalysts.

1 Introduction

As society moves away from fossil fuels and towards renewable energy sources, technological advances such as renewable energy storage and efficient energy production become key drivers in this transition [42, 16]. Since peak energy usage hours differ from energy production hours [32], known as the "duck" curve, energy storage can ensure that cities obtain the energy they need at the right time. However, methods suitable for large-scale storage, such as hydrogen energy storage (HES) [2], require effective catalyst materials both for energy storage (via electrolysis) and for energy retrieval (via fuel cells) [42, 31]. Most catalysts that are currently relied upon to accelerate both energy storage and retrieval are expensive to manufacture [25]. For this reason, the search for efficient, cheap, and durable catalysts is vital to increasing the adoption of renewable energy.

Machine learning, growing in its applications and reaching state-of-the-art performance in image generation [10], language modelling [39], and scientific discovery [41], has been applied to predict catalyst properties and generate new catalysts [33, 13, 8, 37]. Graph neural networks such as FAENet [13, 35, 17] are able to predict adsorption energies, which are a key predictor of catalytic efficiency [29, 27]. Methods such as Crystal-GFN [1] are capable of generating molecules and crystal structures within specified symmetry constraints, which meet certain desirable properties. These machine learning methods can be broadly divided into two categories, predictive models (used for catalyst evaluation) [43, 12], and generative models (used for catalyst design) [33, 28].

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: .

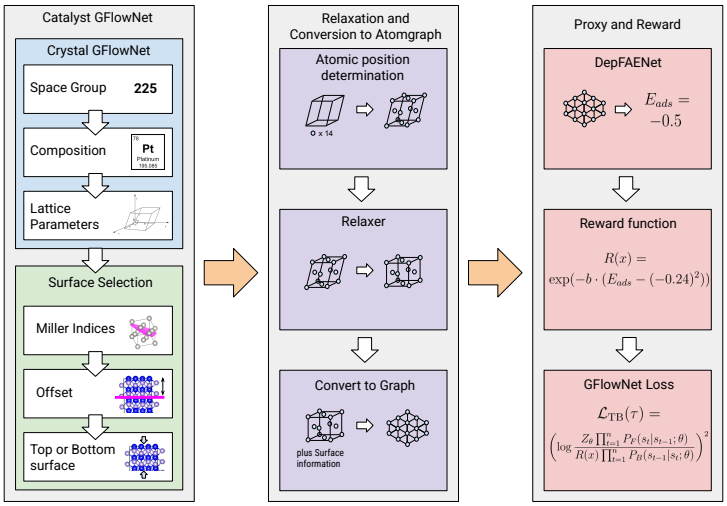

Figure 1: Overview of the Catalyst GFlowNet framework. The leftmost section, Catalyst GFlowNet, samples the catalyst surface. The middle section determines atom positions, relaxes the structure, and converts to a graph. The rightmost section obtains the adsorption energy from a predictive model and the reward function to train the GFlowNet.

Here, we introduce Catalyst GFlowNet (Figure 1), a novel catalyst discovery method building on Crystal-GFN [1], designed to generate a diverse set of efficient catalyst materials. Generative flow networks (GFlowNets or GFNs) are generative models that are able to sample sequentially constructed objects from a desired distribution. This property allows our framework to propose a variety of catalyst materials rather than a single optimal material. Subsequent real-world experiments have a greater chance of discovering a truly high-quality catalyst material when they are able to test multiple proposed materials. Our approach for catalyst design builds on crystal generation by constructing a periodic crystal and then cutting a surface that acts as the catalyst. Additionally, we integrate ML-based structure relaxations into our framework to ensure that the final samples are relaxed (stable). As a signal on the quality of the catalyst material, we use a Graph Neural Network (GNN) based on FAENet [13] to estimate the adsorption energy for a given adsorbate molecule in a reaction of interest. A tailored reward function ensures that we generate realistic crystals that can then serve as the basis for efficient catalysts. In our case study, we verify that Catalyst GFlowNet is able to rediscover the best-known catalysts, demonstrating its promising value for catalyst design.

2 Methodology

In this section, we introduce the general architecture and workflow of the Catalyst GFlowNet (Section 2.1) and the experimental setup for the hydrogen evolution reaction (HER) case study (Section 2.2). Generally, a GFlowNet learns to sample objects proportionally to their reward; we leverage this property while imposing constraints on the way that samples are constructed. Background material on GFlowNets and catalysts for HER is covered in the appendix, Section A.

2.1 Architecture

The Catalyst GFlowNet builds on the Crystal-GFN framework. The overall framework, depicted in Figure 1, consists of three sequential steps: sample generation, sample preparation, and reward computation. After the reward is computed, a GFlowNet objective, such as the trajectory balance loss [26], can be used to train the GFlowNet to sample catalyst representations proportionally to the reward function.

In the first phase, the catalyst surface is sequentially constructed by the GFlowNet. Here, we Crystal-GFN to build catalyst surfaces given a crystal structure. In particular, this involves first selecting the space group (integer between 1 and 230, possibly decomposed into lattice system and point group); following this, selecting the composition (e.g. $A_xB_yC_z$); and finally, selecting the lattice

parameters (lattice vectors a, b, c and angles between them α, β, γ). Conditioned on this crystal representation, the Catalyst GFlowNet selects the Miller indices [40] (three integers ranging from -2 to 2) representing the surface of the periodic crystal object that would act as the catalyst. For example, the Miller indices (100) cut a plane that is parallel to the Y-Z plane created by the lattice vectors[1]. Following this, the offset for the surface plane is selected (continuous value between 0 and 1), which represents the location of the cut in the direction perpendicular to the Miller plane. Finally, the network selects whether the top or bottom surface (represented by a boolean) is used. For a single-element material, such as platinum, the offset and top/bottom surface boolean do not change the resulting surface. However, for more complex materials, such as oxides, the offset and the surface boolean can lead to surfaces with widely different properties.

In the second phase (middle part of Figure 1), we obtain atomic positions symmetrically compatible with the crystal representation by using PyXtal's structure generator [15]. After the structure is fully determined, M3GNet [9] is used to relax the atoms and the unit cell. Then, a converter module creates a surface object using the final structure and sampled Miller indices, top surface information, and offset. Finally, this surface object is converted into a graph, where each node represents an atom. The adsorbate is its own graph, disconnected from the surface [7].

In the final phase (right-most side of Figure 1), the atom graph is passed to the proxy model, DepFAENet in our case, which is a GNN based on FAENet [13], equivariant to Euclidean symmetries, trained to predict adsorption energies for different adsorbates. DepFAENet differs from the original FAENet implementation in that the surface and adsorbate are disconnected graphs [7]. The model predicts the adsorption energy directly without situating the adsorbate on the surface. Then, the predicted adsorption energy is passed to the reward function (described in detail in Section 2.2). The rewards obtained for the constructed samples are then used to compute the trajectory balance loss to train the GFlowNet.

2.2 Case study: Hydrogen Evolution Reaction

For this case study, our aim is to generate stable catalysts for the hydrogen evolution reaction (HER) that have the smallest possible *overpotential*. In catalysis, the overpotential (η) is the amount of electric potential that needs to be applied on top of the theoretical reaction potential; it is a measure of the efficiency of a chemical reaction and its catalyst. For HER, in acidic conditions, it can be predicted directly from the adsorption free energy of hydrogen [29] (which is what our GNN proxy is trained to predict) plus a small correction. Here, to obtain proposed catalysts (x) with low overpotentials, we use the following reward function:

$$R(x) = \exp(-b\eta^2) \quad \text{where} \quad \eta = E_H(x) + E_{corr} \tag{1}$$

We set $b = 100$ and the Gibbs correction $E_{corr} = -0.24$. Although the reward only takes the overpotential (η) into account, structure relaxation for each sample (middle component of Figure 1) ensures that we generate reasonably stable structures. Additionally, at sampling time, structures that have high formation energy among structures with the same composition are discarded.

As a proof of concept, we test our workflow on a highly restricted search space, where atomic positions are fully determined given the space group and the number of atoms. In this search space, we have an expectation of what materials should be sampled the most. With other methods, such as those based on large language models [28], imposing constraints on crystal symmetry, space group, and number of atoms would be much more difficult, if not impossible. The search space that we consider matches closely the catalysts described in Nørskov et al. [29]. The full description of the experimental setting can be found in Table 1 in the appendix. We consider 12 possible elements, including platinum, the best known HER catalyst [29]. We consider the space of single-element structures only, as per Nørskov et al. [29]. Taking the most common form of each structure as found in the Materials Project [21], we note the space group of each structure and restrict the search space to those space groups only. These are space groups $Fm\overline{3}m$ (225) and $Im\overline{3}m$ (229).

[1]Miller indices hkl denote the 3D plane (and those parallel to it) that intersects the x-axis at 1/h, the y-axis at 1/k, the z-axis at 1/l.

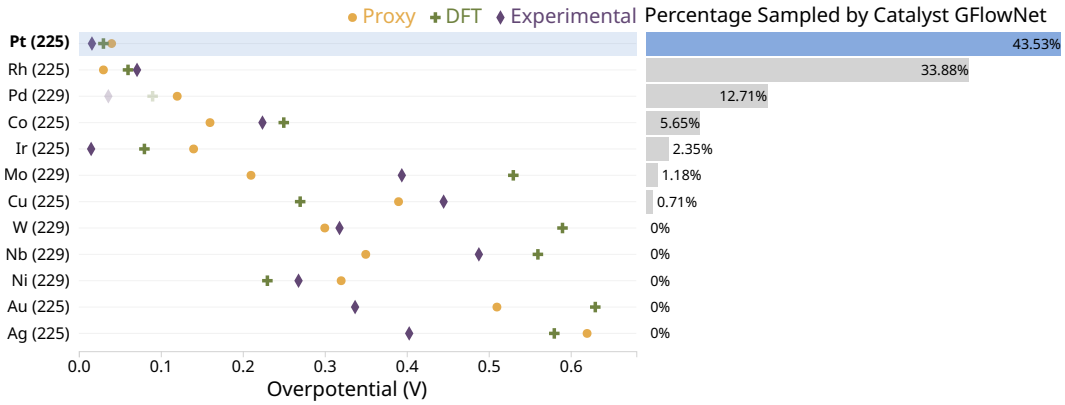

Figure 2: Proportion of sampled structures (right) for the hydrogen evolution reaction case study. Low overpotential (left) is a predictor for efficient catalysts and aligns with the higher sampling rates. The values are detailed in Table 2 in the appendix. Pd(229) does not have experimental or DFT overpotentials, corresponding values for Pd(225) are displayed with shaded colors for reference.

3 Results

In this section, we analyze the catalysts sampled by the Catalyst GFlowNet after being trained, following the experimental setup for the hydrogen evolution reaction (HER). We sampled 1000 structures and performed a M3GNet-based relaxation on the crystal lattice. Then, we filter out all structures that relax to a high formation energy. Specifically, we compute the minimal sampled formation energy per composition, and keep only those relaxed structures that have a formation energy within 0.05 eV of the minimum energy. We do this because in some cases the relaxation strays too far from the true basin given the starting lattice parameters, and as such would not be a stable structure. Then, for each composition, we keep only the space group that has the lowest minimum energy, since the one with the higher energy is estimated to be unstable. Finally, we cut the surface according to the parameters chosen by the Catalyst GFlowNet to construct the final catalyst surface.

The filtered set of relaxed structures contains 425 structures. Figure 2 details the percentage of samples of each of the structures in the paper. As expected, the most frequently sampled structures are those that have a low overpotential, as predicted by the proxy model and DFT values by Nørskov et al. [29] (see Figure 2, left). Platinum (Pt) is known to be the best catalyst for HER, but rhodium (Rh) also performs well, as evidenced by experimental, DFT, and proxy values. We find in fact that rhodium and platinum are the most sampled structures. All generated structures except palladium with space group 229 are found in Nørskov et al. [29] (and are experimentally observed [21]). Palladium with space group 225, the one that is experimentally observed, was sampled, but filtered out because its M3GNet-predicted formation energy was slightly higher than its 229 counterpart. Overall, this experiment validates that the Catalyst GFlowNet framework can independently rediscover the best known HER catalysts, and sample them proportionally to their performance.

4 Discussions and Conclusion

In this paper, we propose a novel catalyst discovery framework that generates complete and relaxed crystal structure descriptions (including atomic positions) for the purpose of finding an inexpensive, efficient catalyst for renewable energy storage. We show that the method can find the best currently known catalyst for the hydrogen evolution reaction (i.e., platinum). We also show that our method samples diverse structures, increasing the likelihood of finding suitable catalysts that withstand real-world testing. Finally, we generate sample structures that our model considers to be good catalysts for the hydrogen evolution reaction; further investigation could shed light on whether these structures could perform well experimentally. This framework is the prototype of a novel method to generate *de novo* stable catalysts for a variety of materials science tasks; integration with real-world experiment and active learning approaches would surely have an impact in the search to reduce the cost of renewable energy storage. In future work, we aim to extend our method in several ways. We aim to test the framework on the oxygen evolution reaction, for which there is currently no efficient,

inexpensive catalyst. Additionally, we aim to add functionality to the Catalyst GFlowNet in order to sample atomic positions. We foresee the extension of this work by embedding the GFlowNet in a multi-fidelity active learning loop [19].

References

[1] Mila AI4Science, Alex Hernandez-Garcia, Alexandre Duval, Alexandra Volokhova, Yoshua Bengio, Divya Sharma, Pierre Luc Carrier, Yasmine Benabed, Michał Koziarski, and Victor Schmidt. Crystal-GFN: Sampling crystals with desirable properties and constraints. *arXiv preprint arXiv:2310.04925*, 2023.

[2] A.Z. Arsad, M.A. Hannan, Ali Q. Al-Shetwi, M. Mansur, K.M. Muttaqi, Z.Y. Dong, and F. Blaabjerg. Hydrogen energy storage integrated hybrid renewable energy systems: A review analysis for future research directions. *International Journal of Hydrogen Energy*, 47(39):17285–17312, 2022.

[3] Lazar Atanackovic, Alexander Tong, Bo Wang, Leo J Lee, Yoshua Bengio, and Jason S Hartford. Dyngfn: Towards bayesian inference of gene regulatory networks with gflownets. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 74410–74428. Curran Associates, Inc., 2023.

[4] Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, and Yoshua Bengio. Flow network based generative models for non-iterative diverse candidate generation. *Advances in Neural Information Processing Systems*, 34:27381–27394, 2021.

[5] Yoshua Bengio, Salem Lahlou, Tristan Deleu, Edward J. Hu, Mo Tiwari, and Emmanuel Bengio. Gflownet foundations. *Journal of Machine Learning Research*, 24(210):1–55, 2023.

[6] Shannon W. Boettcher. Introduction to green hydrogen. *Chemical Reviews*, 124(23):13095–13098, 2024. PMID: 39659176.

[7] Alvaro Carbonero, Alexandre Duval, Victor Schmidt, Santiago Miret, Alex Hernandez-Garcia, Yoshua Bengio, and David Rolnick. On the importance of catalyst-adsorbate 3d interactions for relaxed energy predictions. *arXiv preprint arXiv:2310.06682*, 2023.

[8] Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, Aini Palizhati, Anuroop Sriram, Brandon Wood, Junwoong Yoon, Devi Parikh, C. Lawrence Zitnick, and Zachary Ulissi. Open catalyst 2020 (oc20) dataset and community challenges. *ACS Catalysis*, 11(10):6059–6072, 2021.

[9] Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 2022.

[10] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10850–10869, 2023.

[11] N Danilovic, Ram Subbaraman, D Strmcnik, Kee-Chul Chang, AP Paulikas, VR Stamenkovic, and Nenad M Markovic. Enhancing the alkaline hydrogen evolution reaction activity through the bifunctionality of ni (oh) 2/metal catalysts. *Angewandte Chemie International Edition*, 51(50):12495–12498, 2012.

[12] Alexandre Duval, Simon V Mathis, Chaitanya K Joshi, Victor Schmidt, Santiago Miret, Fragkiskos D Malliaros, Taco Cohen, Pietro Lio, Yoshua Bengio, and Michael Bronstein. A hitchhiker's guide to geometric gnns for 3d atomic systems. *arXiv preprint arXiv:2312.07511*, 2023.

[13] Alexandre Agm Duval, Victor Schmidt, Alex Hernández-García, Santiago Miret, Fragkiskos D. Malliaros, Yoshua Bengio, and David Rolnick. FAENet: Frame averaging equivariant GNN for materials modeling. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt,

Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 9013–9033. PMLR, 23–29 Jul 2023.

[14] Lixin Fan, Zhengkai Tu, and Siew Hwa Chan. Recent development of hydrogen and fuel cell technologies: A review. *Energy Reports*, 7:8421–8446, 2021.

[15] Scott Fredericks, Kevin Parrish, Dean Sayre, and Qiang Zhu. Pyxtal: A python library for crystal structure generation and symmetry analysis. *Computer Physics Communications*, 261:107810, 2021.

[16] Sergio Freeman and Ertan Agar. The impact of energy storage on the reliability of wind and solar power in new england. *Heliyon*, 10(6), 2024.

[17] Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems*, 34:6790–6802, 2021.

[18] Geoffroy Hautier, Anubhav Jain, and Shyue Ping Ong. From the computer to the laboratory: materials discovery and design using first-principles calculations. *Journal of Materials Science*, 47:7317–7340, 2012.

[19] Alex Hernández-García, Nikita Saxena, Moksh Jain, Cheng-Hao Liu, and Yoshua Bengio. Multi-fidelity active learning with GFlownets. *Transactions on Machine Learning Research*, 2024. Expert Certification.

[20] Atsushi Ishikawa. Heterogeneous catalyst design by generative adversarial network and first-principles based microkinetics. *Scientific Reports*, 12(1):11657, 2022.

[21] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin a. Persson. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013.

[22] MichałKoziarski, Andrei Rekesh, Dmytro Shevchuk, Almer van der Sloot, Piotr Gaiński, Yoshua Bengio, Cheng-Hao Liu, Mike Tyers, and Robert A. Batey. Rgfn: Synthesizable molecular generation using gflownets. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 46908–46955. Curran Associates, Inc., 2024.

[23] Romain Lacombe, Lucas Hendren, and Khalid El-Awady. Adsorbrl: Deep multi-objective reinforcement learning for inverse catalysts design, 2023.

[24] Nung Siong Lai, Yi Shen Tew, Xialin Zhong, Jun Yin, Jiali Li, Binhang Yan, and Xiaonan Wang. Artificial intelligence (ai) workflow for catalyst design and optimization. *Industrial & Engineering Chemistry Research*, 62(43):17835–17848, 2023.

[25] Taeho Lim and Soo-Kil Kim. Non-precious hydrogen evolution reaction catalysts: Stepping forward to practical polymer electrolyte membrane-based zero-gap water electrolyzers. *Chemical Engineering Journal*, 433:133681, 2022.

[26] Nikolay Malkin, Moksh Jain, Emmanuel Bengio, Chen Sun, and Yoshua Bengio. Trajectory balance: Improved credit assignment in gflownets. *Advances in Neural Information Processing Systems*, 35:5955–5967, 2022.

[27] Isabela C. Man, Hai-Yan Su, Federico Calle-Vallejo, Heine A. Hansen, José I. Martínez, Nilay G. Inoglu, John Kitchin, Thomas F. Jaramillo, Jens K. Nørskov, and Jan Rossmeisl. Universality in oxygen evolution electrocatalysis on oxide surfaces. *ChemCatChem*, 3(7):1159–1165, 2011.

[28] Dong Hyeon Mok and Seoin Back. Generative pretrained transformer for heterogeneous catalysts. *Journal of the American Chemical Society*, 146(49):33712–33722, 2024. PMID: 39576215.

[29] J. K. Nørskov, T. Bligaard, A. Logadottir, J. R. Kitchin, J. G. Chen, S. Pandelov, and U. Stimming. Trends in the exchange current for hydrogen evolution. *Journal of The Electrochemical Society*, 152(3):J23, jan 2005.

[30] Janghoon Ock, Chakradhar Guntuboina, and Amir Barati Farimani. Catalyst energy prediction with catberta: Unveiling feature exploration strategies through large language models. *ACS Catalysis*, 13(24):16032–16044, 2023.

[31] Ahmed I Osman, Neha Mehta, Ahmed M Elgarahy, Mahmoud Hefny, Amer Al-Hinai, Ala'a H Al-Muhtaseb, and David W Rooney. Hydrogen production, storage, utilisation and environmental impacts: a review. *Environmental Chemistry Letters*, pages 1–36, 2022.

[32] Giovani Manuel Pitra and KS Sastry Musti. Duck curve with renewable energies and storage technologies. In *2021 13th International conference on computational intelligence and communication networks (CICN)*, pages 66–71. IEEE, 2021.

[33] Oliver Schilter, Alain Vaucher, Philippe Schwaller, and Teodoro Laino. Designing catalysts with deep generative models and computational data. a case study for suzuki cross coupling reactions. *Digital discovery*, 2(3):728–735, 2023.

[34] Oliver Schilter, Alain Vaucher, Philippe Schwaller, and Teodoro Laino. Designing catalysts with deep generative models and computational data. a case study for suzuki cross coupling reactions. *Digital discovery*, 2(3):728–735, 2023.

[35] Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[36] Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pages 9377–9388. PMLR, 2021.

[37] Richard Tran, Janice Lan, Muhammed Shuaibi, Brandon M Wood, Siddharth Goyal, Abhishek Das, Javier Heras-Domingo, Adeesh Kolluru, Ammar Rizvi, Nima Shoghi, et al. The open catalyst 2022 (oc22) dataset and challenges for oxide electrocatalysts. *ACS Catalysis*, 13(5):3066–3084, 2023.

[38] Sergio Trasatti. Work function, electronegativity, and electrochemical behaviour of metals: Iii. electrolytic hydrogen evolution in acid solutions. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, 39(1):163–184, 1972.

[39] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[40] Klaus Wandelt. *Encyclopedia of interfacial chemistry: surface science and electrochemistry*. Elsevier, 2018.

[41] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

[42] C Lawrence Zitnick, Lowik Chanussot, Abhishek Das, Siddharth Goyal, Javier Heras-Domingo, Caleb Ho, Weihua Hu, Thibaut Lavril, Aini Palizhati, Morgane Riviere, et al. An introduction to electrocatalyst design using machine learning for renewable energy storage. *arXiv preprint arXiv:2010.09435*, 2020.

[43] C Lawrence Zitnick, Lowik Chanussot, Abhishek Das, Siddharth Goyal, Javier Heras-Domingo, Caleb Ho, Weihua Hu, Thibaut Lavril, Aini Palizhati, Morgane Riviere, et al. An introduction to electrocatalyst design using machine learning for renewable energy storage. *arXiv preprint arXiv:2010.09435*, 2020.

[44] Hao Zou, Haochen Zhao, Mingming Lu, Jiong Wang, Zeyu Deng, and Jianxin Wang. Predicting thermodynamic stability of inorganic compounds using ensemble machine learning based on electron configuration. *Nature Communications*, 16(1):203, 2025.

A Background

In this section, we describe Generative Flow Networks in detail (Section A.2), and we perform a survey of related work that tackles the catalyst discovery problem (Section A.3).

A.1 Catalysts for hydrogen energy storage

Hydrogen Energy Storage (HES) is a scalable, sustainable option for storing excess energy for later use. Although HES has a relatively low efficiency compared to alternatives such as pumped storage hydroelectricity and battery storage [42, 31, 6], it is more scalable and transportable than the alternatives. HES via water electrolysis (water-splitting) can be thought of as two half-cell reactions: the hydrogen evolution reaction and the oxygen evolution reaction. Both HER and OER are electrochemical reactions which occur in the catalyzer at different electrodes. While HER is a one-step reaction, OER consists of four intermediate reactions which, under an applied voltage, produces oxygen gas, protons and electrons under acidic conditions. HER, in this setting, brings electrons and protons together to produce hydrogen gas. As the energy is used for water electrolysis, the hydrogen gas can be collected and stored in tanks, or other storage containers. When the energy is in demand again, fuel cells can extract the energy from the stored hydrogen at about 60% efficiency [14]. An efficient catalyst, used in the fuel cell or in the electrocatalyzer, can significantly speed up the reaction rate and improve the efficiency of the process [42].

A.2 Generative Flow Networks

A generative flow network (GFlowNets, GFN) is an inference method introduced by [4] that allows one to sequentially construct objects with probability proportional to a non-negative reward function. They have been employed in tasks such as molecule generation, crystal generation, and for causal discovery [22, 1, 3]. A desirable capability of GFlowNets is their ability to generate a diversity of objects that all have high reward (relative to other possible candidates in the space). Within catalyst discovery, this is key, as the reward function may not capture all the relevant predictors of whether a material will be a good catalyst. For example, the reward function may capture a target adsorption energy for the catalyst, but not its stability, which is an important trait (high stability) if the catalyst is to be used in real-world applications. After generating a variety of high-reward catalyst materials, one or more of these materials may also be found to be stable during experimental testing. Furthermore, GFlowNets perform particularly well in cases where the search space is quite large, and outperform methods such as MCMC and reinforcement learning in terms of mode mixing [5].

A key component of GFlownets is the proxy model, which is how the GFlowNet obtains information on the reward landscape, and hence what "desirable", high-reward samples are. Proxy models in our context are predictive models (possibly machine learning models) that allow one to evaluate the reward of a particular sampled object. In the case of molecule or crystal discovery, it may be that the sampled molecule or crystal does not exist in any database, and as such, relevant properties of the crystal must be computed in real time instead. Here, we note that for certain properties such as the formation energy of a crystal, physics simulations such as density functional theory (DFT) are traditionally performed. These simulations are computationally expensive and not scalable to high batch sizes. Hence, a machine learning model that is able to replace the physics simulation and compute the properties of interest is valuable; for crystals, there are many works that develop machine learning-based approaches to compute properties such as formation energy and adsorption energy [35, 36, 13, 9].

Crystal GFlowNet is a GFlowNet designed to discover new materials by sequentially sampling crystal structures [1]. This work uses the GFlowNet paradigm to construct and sample crystal structures that have a high reward. The authors test the method on a reward based upon formation energy, where a high reward corresponds to a low formation energy, and a low reward corresponds to a high formation energy. This allows the framework to be used to sample thermodynamically stable materials with desirable properties. As mentioned earlier, the reward function may use any measurable (or predicted, via a proxy model) characteristic of the material. Sequential crystal construction proceeds in three steps. First, the space group (symmetry operations) of the crystal are selected. Then, the composition (elements and their ratios), and finally, the lattice parameters, which determine the shape of the final lattice. Atom positions are not sampled by the model.

A.3 Machine learning for catalyst discovery

Since traditional catalyst design methods are time-consuming and involve trial and error experiments within a very large search space [18], scientists are turning to machine learning approaches to more efficiently explore the space of good catalysts. Recent work has leveraged both predictive models (which evaluate the performance of catalysts based on their measurable attributes) and generative models (which aim to construct catalyst representations with desirable properties) for catalyst discovery.

One major thread of predictive machine learning for catalyst design are predictive models to replace density functional theory (DFT) calculations of adsorption energies. The Open Catalyst Dataset 2020 [8] and 2022 [37] are datasets of 1.2 million and 62 thousand relaxations respectively. The datasets span a wide variety of materials, surfaces and adsorbates, with the inclusion of oxides in OC22. Machine learning models such as FAENet, Schnet, PaiNN, and Graphormer [13] are neural networks (e.g. graph neural networks, transformers) that can be trained to perform the OC20 Initial Structure to Relaxed Energy (IS2RE) task, which involves predicting the relaxed energy of an adsorbate-catalyst system from their initial atomic positions. Machine learning models have also been used to predict formation energies, perform relaxations, predict forces [9]. Predictive machine learning models have also been used to predict the thermodynamic stability of catalysts [44]. All of these predictive models have the potential to be included as proxy models within generative approaches.

Recent generative approaches to catalyst discovery include reinforcement learning [23], large language models [24], variational autoencoders [34], and generative adversarial networks [20] which aim to generate catalysts *de novo*, by meeting criteria such as target adsorption energies and encoding this in the reward or loss. AdsorbRL [23] employs Deep Q-learning to find compositions of catalysts that have either minimal or maximal adsorption energy. They only model composition, omitting space groups, crystal symmetries, surface selection, and atom positions. [34] addresses the problem by employing a VAE to generate catalysts by sampling from the latent space, but we note that reliance on the latent space makes controllable generation and systematic exploration of possible catalysts difficult. [20] takes the approach of GANs to generate catalysts for the ammonia formation reaction, but does not take catalyst stability into account during generation. LLM-based tools such as [24, 30] have so far been mainly applied to help screen known catalysts and perform a literature search.

B Effect of relaxation on lattice parameters

After the GFlowNet generates samples, we relax them using M3GNet PES [9], as detailed in Section 2.1. Figure 3 demonstrates the change in lattice parameters of the cubic platinum lattice before and after relaxation. We note that the lattice parameters relax to the minimum of the total energy of the system.

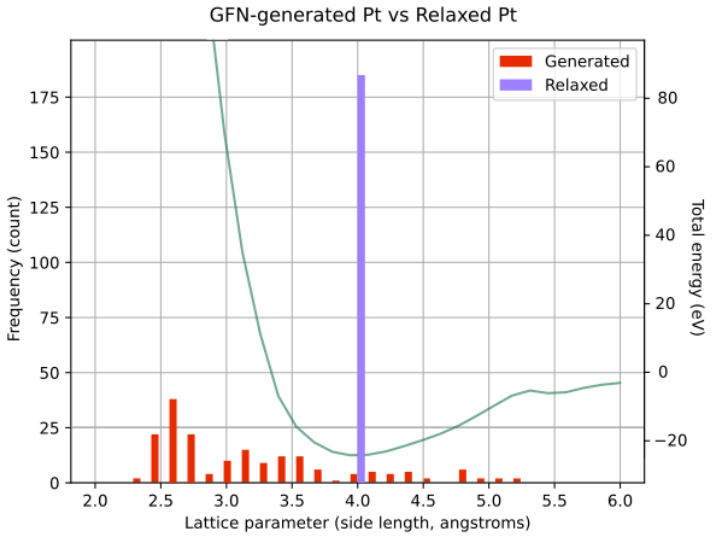

Figure 3: For the HER case study, the platinum samples generated by the GFlownet relax to the minimum energy every time. The green line represents the total energy of the system.

Table 1 details the search space for the Hydrogen evolution reaction case study (Section 2.2). We note that in general, the Catalyst GFlowNet's constraints and search space can be set to be much larger than that used in the HER case study. In future work, we will consider a larger search space up to 80-100 atoms, more elements and space group options, while enforcing a neutral charge constraint.

Table 1: Hydrogen evolution reaction catalyst search space.

Setting	Value	Units
Possible elements	Pt, Ag, Au, Pd, Ir, Ni, W, Co, Cu, Mo, Rh, Nb	
Min. # of different elements in unit cell	1	
Max. # of different elements in unit cell	1	
Min. atoms in unit cell	2	
Max. atoms in unit cell	4	
Min. atoms per element	2	
Max. atoms per element	4	
Enforce neutral charge?	No	
Possible space groups	225, 229	
Min. lattice parameter length	2	Angstroms
Max. lattice parameter length	6	Angstroms
Min. angle between lattice vectors	60	Degrees
Max. angle between lattice vectors	140	Degrees

C Table of hydrogen evolution reaction samples

Table 2: Proportions of sampled structures for the case study: hydrogen evolution reaction.

Composition	Space Group	Overpotential (eV)			Samples	
		Experimental[2]	DFT	Proxy Model	Count	Percentage
Pt	225	0.016	0.03	0.04	185	**43.53**
Rh	225	0.071*	0.06	0.03	144	33.88
Pd	229	0.036*	0.09*	0.12	54	12.71
Co	225	0.224*	0.25	0.16	24	5.65
Ir	225	0.015	0.08	0.14	10	2.35
Mo	229	0.394*	0.53	0.21	5	1.18
Cu	225	0.445	0.27	0.39	3	0.71
W	229	0.318*	0.59	0.30	0	0.00
Nb	229	0.488*	0.56	0.35	0	0.00
Ni	229	0.268	0.23	0.32	0	0.00
Au	225	0.337	0.63	0.51	0	0.00
Ag	225	0.403	0.58	0.62	0	0.00
Sum					425	100.00

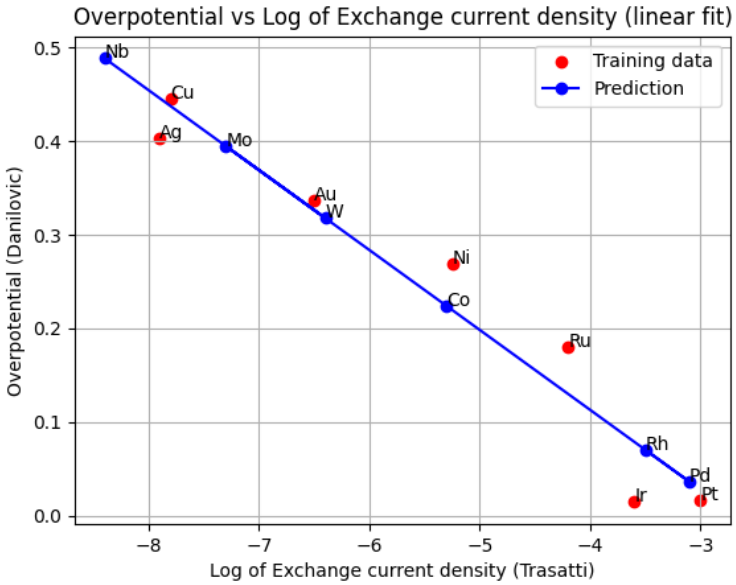

Figure 4: A linear fit that maps log of exchange current density to experimental overpotential using data from [38] and [11]. These values are used column 3 (from the left) of Table 2.

[2]When available, the experimental overpotential was added from [11]. For the remaining elements, marked with an asterisk, the overpotentials were inferred using a linear fit between the log of experimental exchange current density (from [38]) and the available experimental overpotentials (from [11]). Figure 4 shows the linear fit and predictions.