

PAPER • OPEN ACCESS

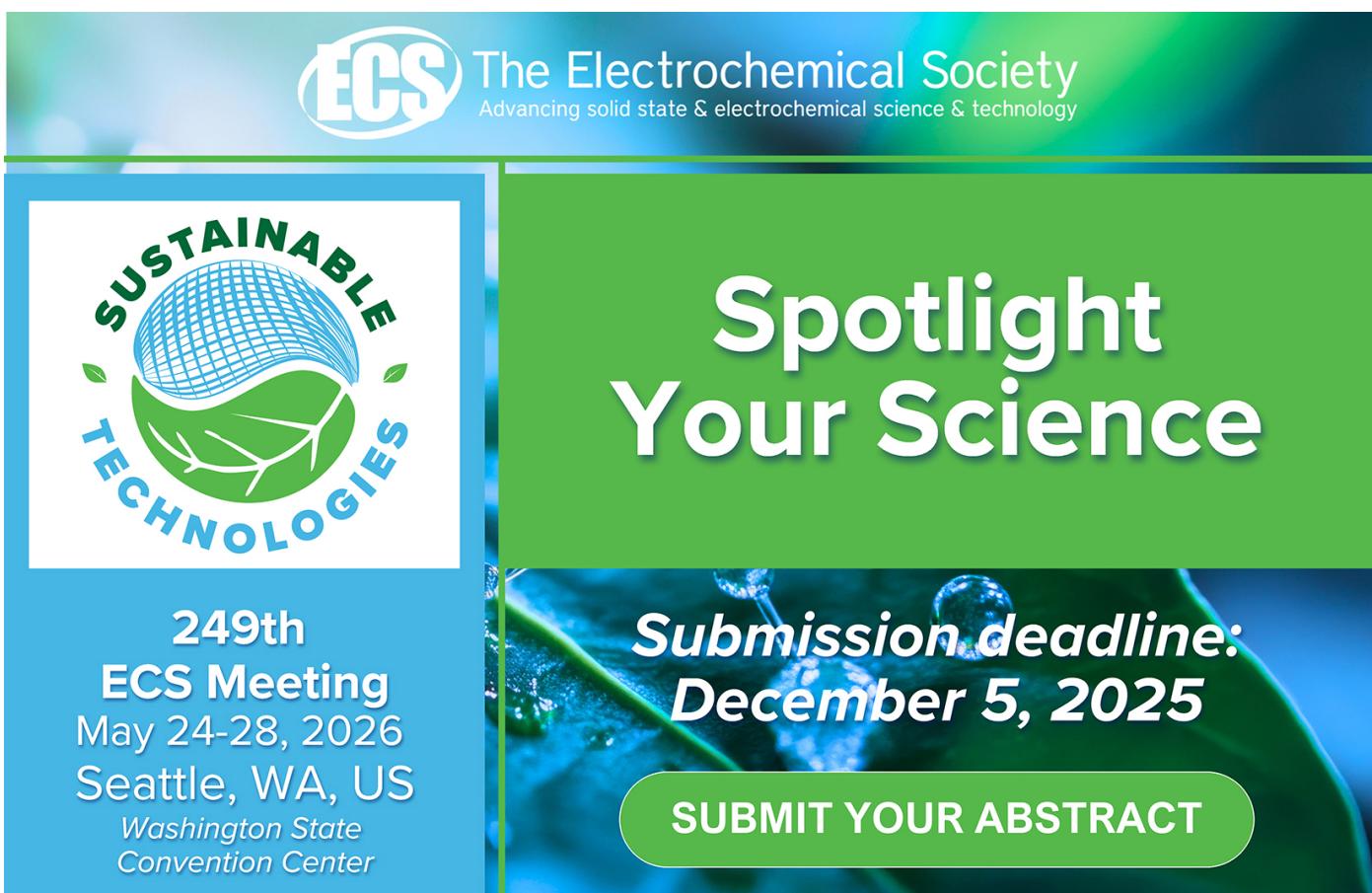
The search for high-entropy fuel-cell catalysts using disorder descriptors

To cite this article: Guangshuai Han *et al* 2025 *Nano Futures* **9** 045001

View the [article online](#) for updates and enhancements.

You may also like

- [pathfinder: A Semantic Framework for Literature Review and Knowledge Discovery in Astronomy](#)
Kartheik G. Iyer, Mikaeel Yunus, Charles O'Neill et al.
- [Real-space orbital tiling approach for the classification and design of superconductors](#)
Gregory Bassan, Wyatt Bunstine, Rebecca Han et al.
- [Photoacoustic imaging of a human vertebra: implications for guiding spinal fusion surgeries](#)
Joshua Shubert and Muyinatu A Lediju Bell



ECS The Electrochemical Society
Advancing solid state & electrochemical science & technology

SUSTAINABLE TECHNOLOGIES

249th ECS Meeting
May 24-28, 2026
Seattle, WA, US
Washington State Convention Center

Spotlight Your Science

Submission deadline:
December 5, 2025

SUBMIT YOUR ABSTRACT

**PAPER****OPEN ACCESS****RECEIVED**
11 July 2025**REVISED**
25 September 2025**ACCEPTED FOR PUBLICATION**
30 October 2025**PUBLISHED**
21 November 2025

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



The search for high-entropy fuel-cell catalysts using disorder descriptors

Guangshuai Han^{1,2,4} , Tianhao Li^{1,4} , Xiao Xu¹ , Jaehyung Lee¹ , Guotao Qiu¹, Sabrina Sequeira¹ , Akshaya Ajith¹ and Corey Oses^{1,2,3,*}

¹ Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, MD 21218, United States of America

² Ralph O'Connor Sustainable Energy Institute, Johns Hopkins University, Baltimore, MD 21218, United States of America

³ William H. Miller III Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, United States of America

⁴ These authors contributed equally to this work.

* Author to whom any correspondence should be addressed.

E-mail: corey@jhu.edu

Keywords: high-entropy alloys, fuel-cell catalysts, materials informatics, machine learning, hydrogen

Abstract

The transition to a hydrogen economy depends on efficient, affordable catalysts for fuel cells. Platinum—the industry standard for fuel-cell electrodes—is costly and scarce, highlighting the need for practical alternatives. High-entropy alloys offer vast compositional diversity and tunable properties that can mitigate these issues, yet their chemical complexity and configurational disorder have hindered rational discovery. Here, we introduce a data-driven framework that couples machine learning with first-principles disorder descriptors—including the entropy forming ability, disordered enthalpy-entropy descriptor, and electronic-structure similarity metrics to platinum—to predict alloy synthesizability and catalytic performance. These descriptors are applied for the first time in the context of fuel-cell catalyst discovery. The workflow rapidly screens more than 20 000 compositions and identifies several platinum-free candidates that are economically viable, readily scalable, and exhibit promising predicted activity. These results demonstrate that disorder descriptors are reliably predicted by machine learning models and can be effectively integrated into materials-discovery pipelines, accelerating innovation across complex compositional spaces.

Hydrogen fuel cells efficiently convert hydrogen and oxygen into electricity and water, providing clean, reliable power for applications ranging from long-haul transportation to warehouse forklifts and distributed energy systems. Proton-exchange-membrane (PEM) fuel cells are especially attractive because of their rapid start-up and high power density [1]. Their widespread adoption, however, is constrained by the reliance on platinum (Pt) as the benchmark catalyst for the oxygen-reduction reaction (ORR) and hydrogen-oxidation reaction (HOR) [2–4]. Pt's high cost and limited supply account for as much as 42% of the total stack cost [5], creating an urgent need for less expensive yet equally effective catalysts.

Alternative catalyst families—transition-metal-nitrogen-carbon composites, perovskite oxides, and metal-organic frameworks among them [6–8]—have been investigated, but few achieve the combination of activity, durability, and manufacturability required for commercial fuel cells [9–11]. High-entropy alloys (HEAs) have recently emerged as a promising platform. Composed of multiple principal elements in near-equimolar ratios [12–14], HEAs exhibit exceptional chemical stability, vast compositional tunability, and surface electronic structures that can be tailored for catalytic reactions. Their high configurational disorder stabilizes unconventional combinations of elements [12, 15–17], opening a design space orders of magnitude larger than that of traditional alloys. The immense design space creates the possibility of discovering alloys that are not only highly active and durable, but also lower in cost and less susceptible to supply chain constraints, without sacrificing catalytic performance. This opportunity comes with a formidable challenge: the number of possible HEA compositions is so large that conventional experimental screening—or even brute-force first-principles calculations—is impractical. Efficient navigation of this space therefore demands data-driven strategies that can connect composition to catalytic synthesizability and performance with minimal computational cost.

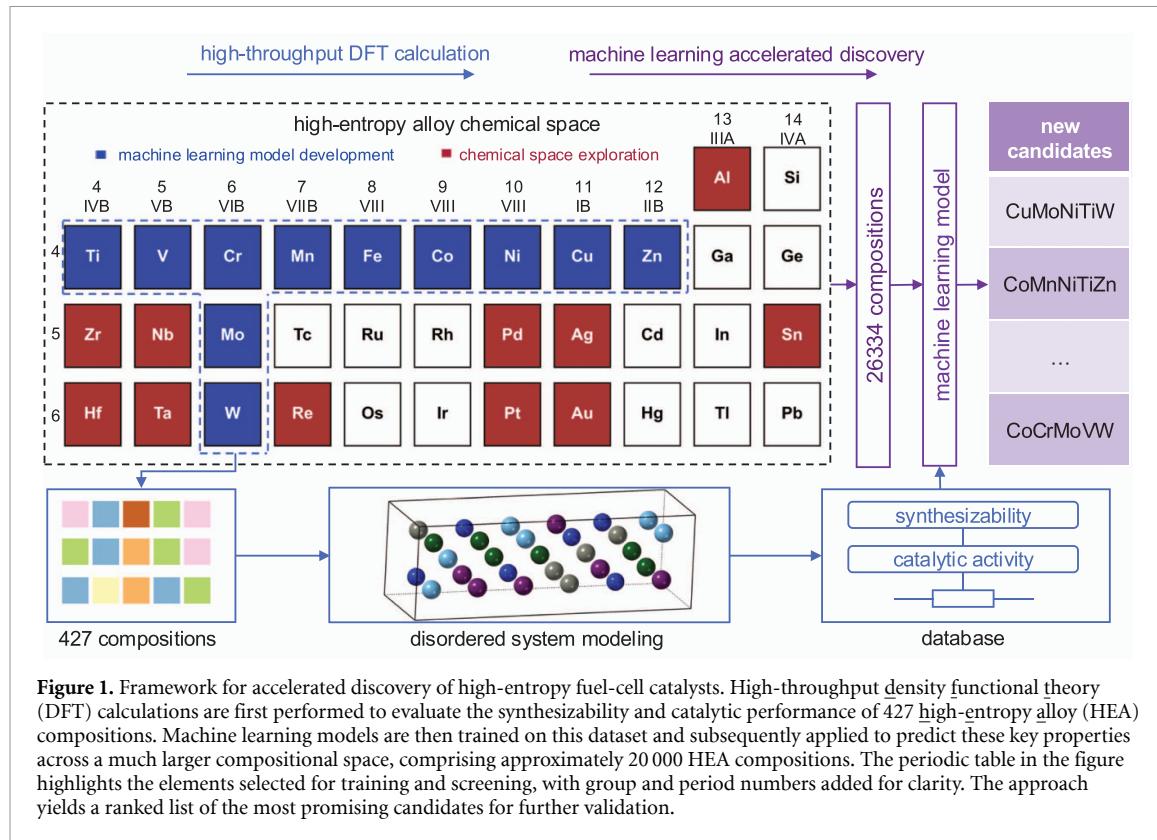


Figure 1. Framework for accelerated discovery of high-entropy fuel-cell catalysts. High-throughput density functional theory (DFT) calculations are first performed to evaluate the synthesizability and catalytic performance of 427 high-entropy alloy (HEA) compositions. Machine learning models are then trained on this dataset and subsequently applied to predict these key properties across a much larger compositional space, comprising approximately 20 000 HEA compositions. The periodic table in the figure highlights the elements selected for training and screening, with group and period numbers added for clarity. The approach yields a ranked list of the most promising candidates for further validation.

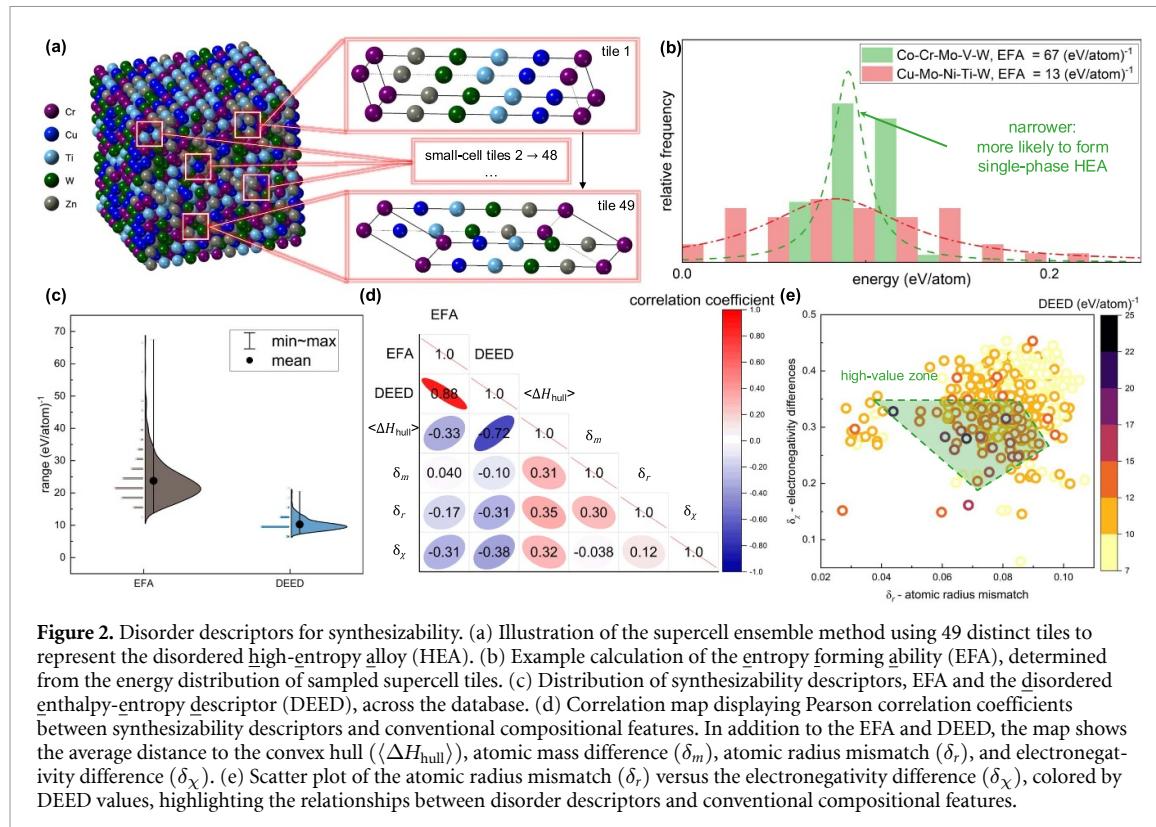
Recent investigations have widely employed machine learning (ML) to support the design of HEAs [15, 18–20]. ML studies on HEAs are typically categorized into three areas [21]: (i) phase formation prediction, which accelerates experimental design by estimating the phase stability for a given composition [22]; (ii) property prediction, where ML models are used to capture specific key properties such as hardness [22], yield strength, and oxidation resistance [23]; and (iii) ML interatomic potentials (MLIPs) [24], often combined with high-throughput calculations to address the challenge of data scarcity.

In parallel, ML is increasingly applied to electrocatalyst design [25]. Descriptor-oriented approaches make use of structural parameters (e.g. coordination numbers, bond lengths), electronic features (such as the *d*-band center and valence electron count), and catalytic descriptors (including adsorption energies of key intermediates) [26]. These descriptors enable data-driven predictions of catalytic activity and stability for various reactions—including hydrogen evolution, oxygen evolution, oxygen reduction, carbon dioxide reduction, and nitrogen reduction—thereby accelerating the discovery of efficient electrode materials [27]. Taken together, these studies illustrate the significant impact of ML approaches on advancing the design of HEAs for electrolysis applications. However, comprehensive disorder descriptors based on validated first-principles approaches—and their systematic integration with ML—remain underexplored for HEAs.

Here we introduce a materials-discovery framework that meets this need (figure 1). We couple high-throughput first-principles calculations with ML to predict the synthesizability and catalytic activity of HEAs. In addition to identifying specific candidate materials, the framework demonstrates how physics-informed disorder descriptors can be systematically incorporated into ML pipelines, thereby accelerating the exploration of complex compositional spaces. Although we focus on fuel-cell catalysts, the same strategy is applicable to other functional materials—ranging from thermoelectrics to structural alloys [13, 28]—thus providing a general pathway toward faster, data-driven materials innovation.

1. Results

Catalyst synthesizability. To model the properties of HEAs, we adopt a supercells-based statistical ensemble strategy [29, 30]. As depicted in figure 2(a), a chemically-disordered system is approximated by a set of symmetrically inequivalent ordered supercells ('supercell tiles' [31, 32]), each sampling a distinct local chemical environment within the full disordered system. The entropy forming ability (EFA) is a descriptor devised within the supercell-ensemble construction quantifying configurational

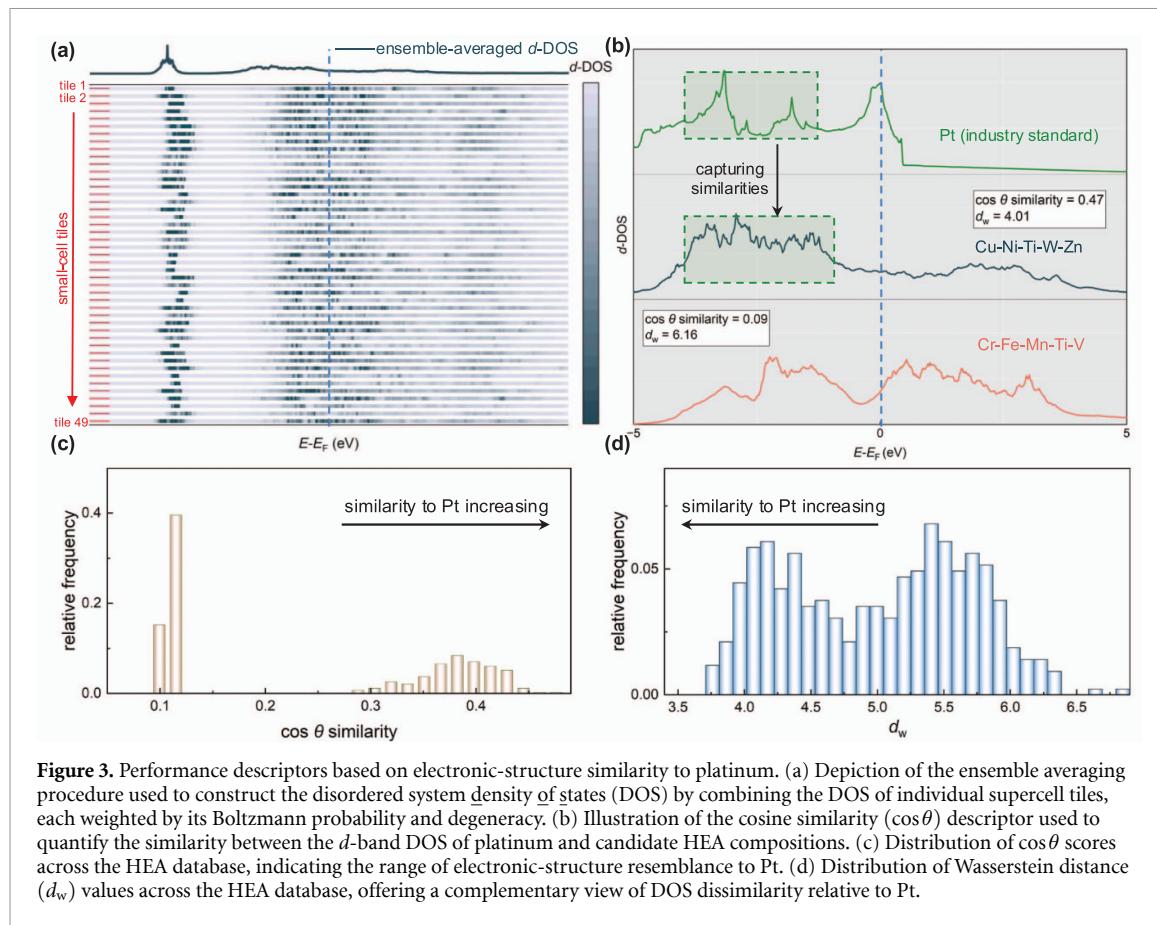


disorder. As illustrated in figure 2(b), $EFA = \sigma \{H_i\}^{-1}$, where $\sigma \{H_i\}$ is the standard deviation of the zero-temperature enthalpies of the supercell tiles [16]. A narrow energy distribution yields a high EFA, signaling that the composition can more readily stabilize a high-entropy single phase. For high-entropy metal carbides, a threshold EFA can be defined, below which systems were shown to phase separate [16], motivating our investigation of EFA as a guiding descriptor for HEAs.

In addition to the EFA, we employ the disordered enthalpy-entropy descriptor [33]: $DEED = \sqrt{EFA/\langle \Delta H_{\text{hull}} \rangle}$, which weighs the entropic gain captured by the EFA against the enthalpic cost of forming a high-entropy system over competing phases—quantified by its distance to the convex hull ($\langle \Delta H_{\text{hull}} \rangle$) [34]. DEED also helps distinguish single-phase high-entropy compositions from those that phase separate and was shown to have applicability to a broader set of chemistries than EFA alone, including metal carbonitrides and metal borides [33]. EFA and DEED both derive from a careful treatment of the thermodynamic density of states [35]. In this framework, the supercells represent ordered approximants of configurationally excited states. See [33] for details.

Figure 2(c) illustrates the distribution of EFA and DEED values across the database. Notably, the DEED distribution is more narrow than that of the EFA, demonstrating how the incorporation of enthalpic contributions creates a more selective pool of viable candidates. To illustrate how DEED compares with other descriptors, figure 2(d) presents a correlation map of key thermodynamic variables and traditional composition-based descriptors—including atomic mass differences among the elements (δ_m), electronegativity differences (δ_X), and atomic radius mismatch (δ_r) [36, 37]. By construction, DEED strongly correlates with EFA and the distance to the hull. Beyond these, the traditional composition-based descriptors display only weak correlations with either EFA or DEED, suggesting limitations on the predictive power of compositional descriptors in resolving the synthesizability of HEAs. Even so, their utility should not be overlooked.

To evaluate whether simple combinations of atomic descriptors can explain trends in synthesizability, we analyze their pairwise correlations. Figure 2(e) plots δ_X vs. δ_r —two descriptors having roots in Hume-Rothery theory for solubility [38]—with DEED values depicted by color. A ‘hot region’ can be distinguished comprising compositions generally exhibiting higher DEED values. A similar but wider region can be resolved based on EFA values. Clusters of high EFA/DEED are not evident from pairwise correlations alone, and suggest opportunities to create analogous Hume-Rothery rules—originally devised for two-component systems—for HEAs.



Catalytic performance. Catalytic performance depends on properties spanning several length scales, including adsorption energies, active site availability, reaction kinetics, and stability/cyclability—making comprehensive modeling challenging. At the atomic scale, the electronic structure determines critical bonding behavior, which cannot be too strong (leading to catalyst poisoning) or too weak (resulting in low reaction rates). Descriptors built on the electronic structure have become essential for rational catalyst design [39]. For example, the d -band model [40] links the position of the d -band center relative to the Fermi level with catalytic activity. A d -band center closer to the Fermi level is a key indicator of lower antibonding-orbital occupation, which enhances the adsorption of oxygenated species (such as O_2 , O^* , and OH^*) during ORR, facilitating critical bond-breaking and bond-forming steps [41–43], and improves interactions with hydrogen during HOR, promoting efficient electron transfer [44]. The electronic structure, including the position of the d -band center, can be tuned through composition [45]. While the d -band model captures important adsorption mechanisms, it can oversimplify factors such as orbital hybridization, local geometry, and the chemical environment that influence catalyst reactivity and selectivity [46, 47].

Instead, we employ two complementary descriptors to quantify the similarity of the full d -band density of states (d -DOS) between HEAs and industry-standard Pt. The d -DOS is represented as a one-dimensional vector across energy, with each element d -DOS_j indicating the number of electronic states between E_j and $E_j + \delta E$. The first descriptor, cosine similarity ($\cos\theta$), measures the general angle between two vectors [48], whereas the second descriptor, the Wasserstein distance (d_w), measures the minimal cost required to transform one DOS distribution into another [49]. The two descriptors capture differences in DOS features (peak intensity, alignment, and energy shifts) distinctly, as highlighted by our subsequent analysis.

Figure 3(a) illustrates the calculation of the d -DOS for disordered HEAs within the supercells framework. Multiple supercell tiles, capturing distinct atomic configurations, contribute individual d -DOS profiles, which are then ensemble-averaged—as a function of temperature—to generate the final disorder-resolved d -DOS. This approach resolves a statistically meaningful representation of the electronic structure in complex disordered systems [16, 50–52].

Examples of the two similarity descriptors are depicted in figure 3(b). The Cu–Ni–Ti–W–Zn alloy yields a high cosine similarity score of 0.47 and a low Wasserstein distance of 4.01, indicating good

alignment in both the shape and energy distribution of its d -orbital DOS with respect to Pt. In contrast, the Cr–Fe–Mn–Ti–V system shows a low cosine similarity of 0.09 and a higher Wasserstein distance of 6.16, reflecting significant electronic dissimilarity. These examples highlight the complementary nature of the two descriptors: $\cos\theta$ captures directional similarity in DOS shape, while d_w accounts for broader discrepancies in peak position and intensity. Together, they provide a more comprehensive assessment of catalytic resemblance to Pt.

The distribution of $\cos\theta$ and d_w scores across the full HEA dataset is shown in figures 3(c)–(d). The majority of candidates cluster at low $\cos\theta$ values, with a smaller tail skewed toward Pt-like systems—reflecting the selectivity of the descriptor and inviting a wider search beyond the initial 427 HEA compositions. Alternatively, the d_w values display a bimodal distribution with broad peaks, echoing the near-continuum of electronic-structure characters accessible with HEAs [53, 54]. While $\cos\theta$ emphasizes local alignment in DOS features, d_w provides a global comparison sensitive to both amplitude variations and energy displacement. The use of both descriptors allows for a tiered-filtering scheme. First, $\cos\theta$ effectively screens highly dissimilar profiles, as evidenced by the concentration of low similarity scores. Then, d_w offers a finer-grained differentiation among the remaining candidates, exhibited by the broad symmetric distribution across the dataset. This more uniform distribution is also expected to facilitate the training of more accurate ML models by providing a balanced and well-resolved label space. These similarity-labeled descriptors serve as training labels within the ML framework, providing domain-informed knowledge that supports model development and screening.

ML modeling. Our objective is to develop ML models with wide applicability domains that enable efficient screening of the vast HEA space. This is achieved by combining traditional composition-based features—derived from the properties of individual elements—with new, validated disorder descriptors such as EFA, DEED, and d -DOS similarity to industry-standard platinum. Together, these inputs help predict catalyst synthesizability and performance, offering a targeted approach for identifying promising candidates within the HEA design space. To accomplish this, we employ a well-established and interpretable ML approach—the random forest regressor [56–59]—to demonstrate that information derived from elemental composition and disorder can reliably predict complex properties in HEAs.

Figure 4 presents the performance of four ML models trained to predict EFA, DEED, d -DOS cosine similarity ($\cos\theta$), and Wasserstein distance (d_w). All models demonstrate strong predictive accuracy, with r^2 testing scores exceeding 0.8 and low mean absolute errors (MAE) across the dataset. Furthermore, five-fold cross-validation yields consistent performance, supporting the generalizability of the models.

The model predicting DEED exhibits the best performance between the synthesizability descriptors. This can be attributed to DEED’s construction as a convex-hull-based descriptor [34], which emphasizes proximity to low-energy competing phases and produces a more regular and compact distribution of values than EFA. By incorporating phase stability considerations, DEED better captures underlying thermodynamics and compositional chemistry, and is thus more compatible with the design of our input features. These statistical characteristics make DEED more amenable to ML than EFA. In contrast, EFA—which measures the standard deviation of energies across different supercell tiles—is more sensitive to energetic outliers. These outliers introduce large label variances and degrade model performance, especially in systems with broad energy landscapes.

Figure 4(c) shows that the model predicting $\cos\theta$ similarity achieves high accuracy, with a testing r^2 exceeding 0.9. This indicates that the model effectively captures the relationship between composition and electronic structure similarity. However, prediction performance in the low-similarity regime (0.1–0.3), which dominates the dataset (as shown in figure 3(c)), remains less accurate due to limited variance in label values. Figure 4(d) presents the model performance for the Wasserstein distance d_w . The d_w values exhibit a more uniform distribution compared to other descriptors, which significantly facilitates model learning. As a result, this model achieves the best performance among all four descriptors, with a testing r^2 of 0.99 and a MAE of 0.06.

Expanded search. We trained ML models to perform high-throughput inference over an expanded chemical composition space. This process, illustrated in figure 5(a), screened over 20 000 five-metal HEA compositions. The element pool, summarized in figure 1, includes additional components commonly used in HEAs [60, 61]: other transition metals, as well as Al and Sn—selected for their size and electronegativity compatibility with transition metals.

We performed multi-objective optimization, as shown in figure 5(a), to identify HEA compositions suitable for experimental validation. Criteria included synthesizability, electronic-structure similarity to Pt, and economic metrics such as cost, recycling rate, abundance, and supply risk. These metrics prioritize candidates for real-world applications and scalable production, promoting manufacturability [62].

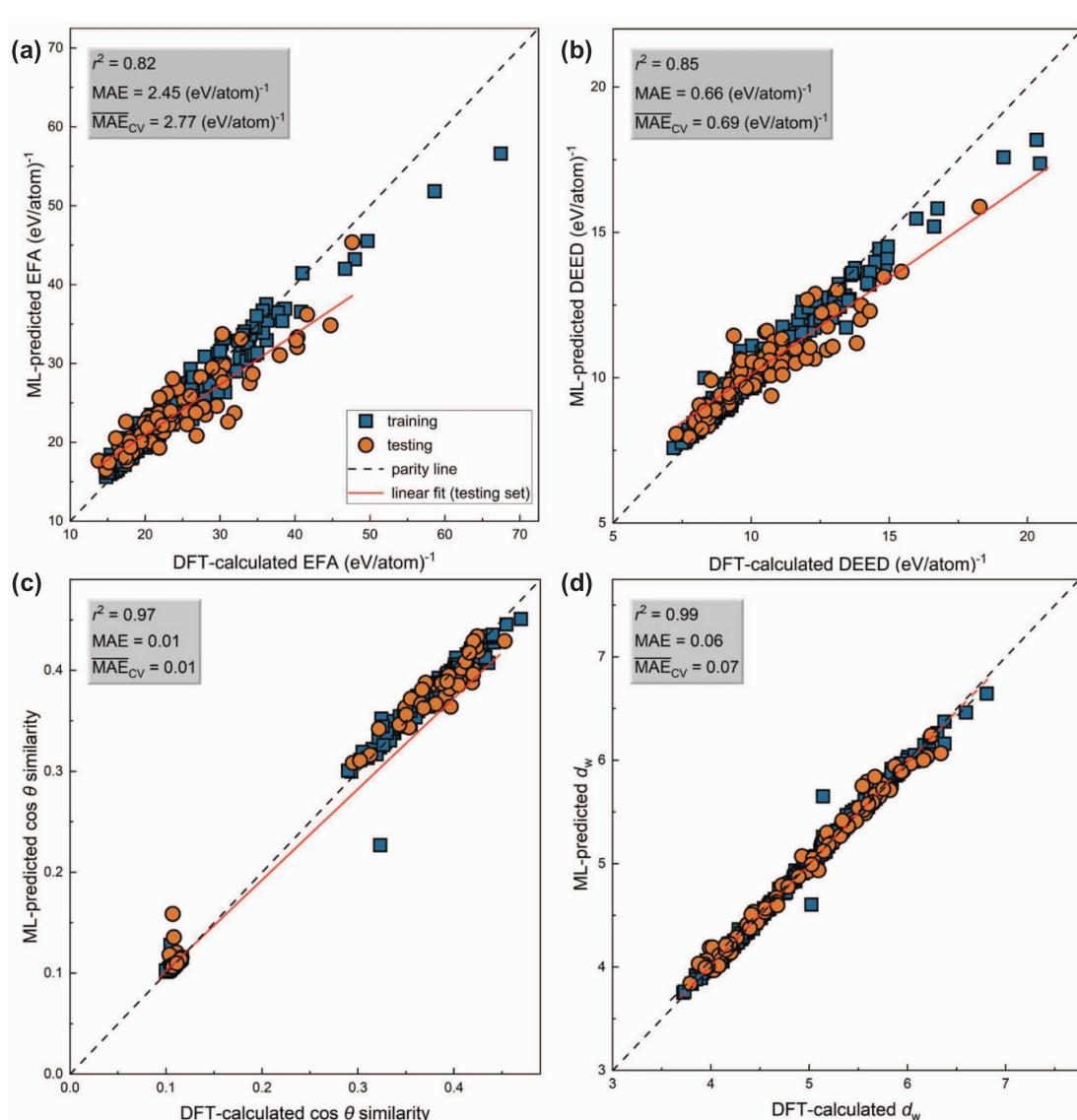
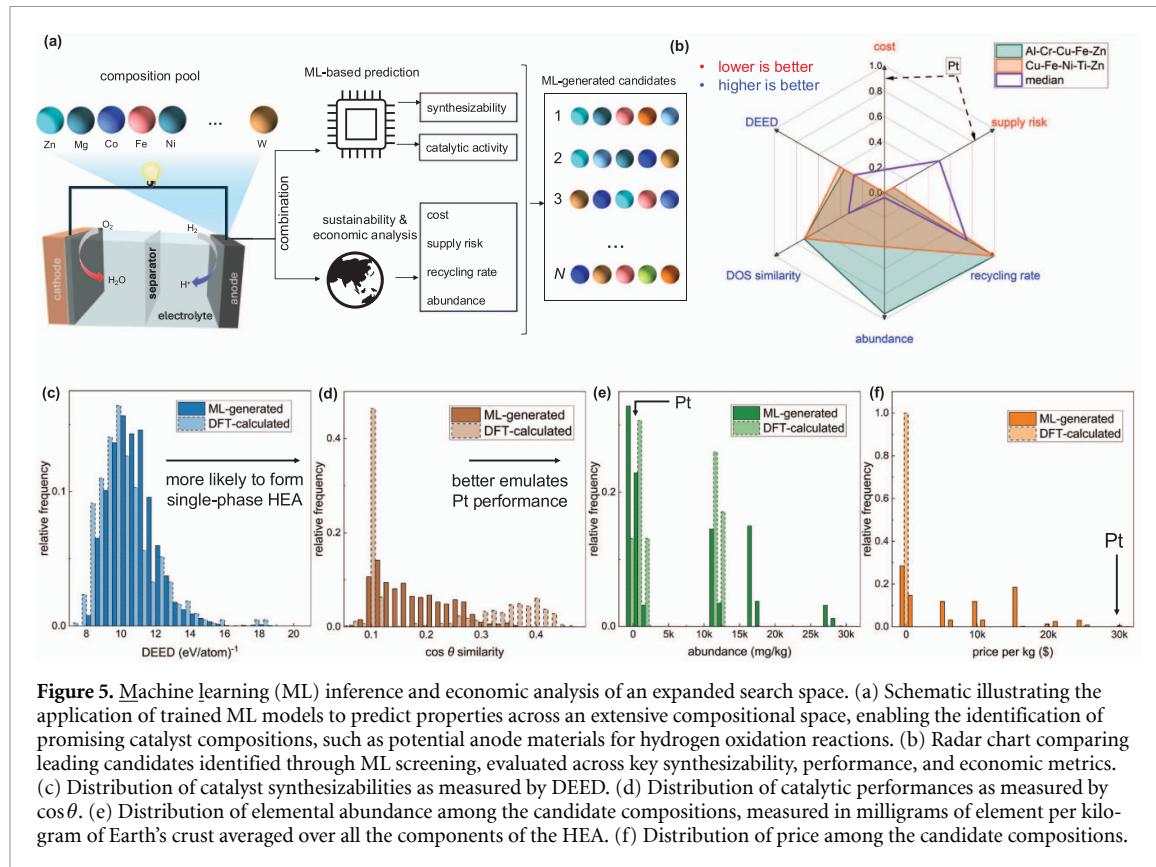


Figure 4. Machine learning (ML) model performance and validation. Parity plots comparing ML-predicted and reference values for four target properties: (a) entropy forming ability (EFA), (b) disordered enthalpy-entropy descriptor (DEED), (c) d -band density of states cosine similarity ($\cos \theta$) to platinum, and (d) Wasserstein distance (d_w). All models exhibit strong predictive accuracy and consistency across training and validation splits. The quantity $\overline{\text{MAE}}_{\text{CV}}$ shown in the legends denotes the average mean absolute error over five-fold cross-validation [55].

For example, figure 5(b) shows a radar chart visualizing the performance of two representative compositions: one selected through ML-based screening (Al–Cr–Cu–Fe–Zn) and another from our original set of DFT calculations (Cu–Fe–Ni–Ti–Zn). The chart compares their relative properties, demonstrating that the ML-discovered alloy matches or exceeds the DFT reference in nearly every category. ML-predicted DEED and $\cos \theta$ distributions, shown in figures 5(c) and (d), reveal trends consistent with our initial DFT dataset and, more importantly, that the range of values predicted by ML is within the range of values provided by our seed calculations. This helps establish that the models are operating within their applicability domain as expected, given the chemical similarity of the additional components in our search. Figure 5(e) highlights several materials composed of elements more readily available than Pt, while figure 5(f) highlights compositions with components significantly lower in cost. Table 1 summarizes the top-performing compositions ranked in the top 25% across all metrics. Among these, 38 originated from ML predictions, while only two came from the original DFT dataset—demonstrating the model’s utility in navigating vast, unexplored spaces to discover catalysts that outperform Pt across multiple critical metrics.



2. Discussion

Computational descriptors. First-principles-based descriptors such as EFA and DEED are derived entirely from quantum-mechanical calculations, using fundamental physical laws to model electronic structure and atomic interactions [64, 65]. Because they are independent of experimental input, EFA and DEED are robust and generalizable across a wide range of material systems—a critical advantage for data-driven discovery, especially in complex or underexplored compositional spaces. These descriptors have successfully predicted the formation of over 60 systems and guided the synthesis of nearly 20 new high-entropy ceramics [16, 33, 58], demonstrating their practical applicability and ability to identify promising disordered materials outside the set of known compounds.

In contrast, widely used semi-empirical approaches such as calculation of phase diagrams (CALPHADs) [66] rely on parameterizing phase Gibbs free energies using thermochemical and phase-equilibria data, primarily from binary and ternary experimental systems. While CALPHAD excels where data is abundant, its reliance on experimental databases limits its predictive power in regions with sparse or incomplete information—a virtual certainty for HEAs due to their immense compositional diversity. As a result, predictions in the quinary HEA space are largely extrapolative, significantly reducing CALPHAD's standalone efficacy for exploratory materials discovery.

However, hybrid approaches that incorporate CALPHAD-derived features alongside first-principles and other key descriptors have proven effective in ML models for predicting formability and stability in high-entropy systems [58, 59]. This demonstrates that combining insights from both approaches can be especially valuable for accelerating discovery and design in complex materials spaces.

Applicability domain. The ML models were trained on 427 quinary HEA compositions (over 20 000 DFT calculations), with elemental coverage spanning all groups and periods of the transition-metal *d*-block. This chemical diversity ensures representation of key trends relevant for HEA screening. The trained models were then applied to predict properties for 20 000 quinary HEA compositions—a 50-fold expansion beyond the initial dataset that would require over a million DFT calculations if performed directly. This substantial expansion warrants a careful evaluation of the models' applicability domain.

The seed DFT dataset enables interpolation within a chemically meaningful space, i.e. with new compositions that are chemically similar to those used during training. ML-predicted property distributions closely match those obtained from seed dataset, and their value ranges do not exceed those observed in

Table 1. Top-ranked HEA candidates based on multiple metrics. Cost denotes the estimated material price in USD per kilogram. Recycling rate represents the proportion of each element that is recovered and reused. Abundance refers to the elemental concentration in the Earth's crust (mg/kg). Supply risk is an index (1–10) that combines abundance, production concentration, substitutability, recycling, and geopolitical factors. All data were obtained from [63]. To generate the ranking, each metric was normalized to [0,1], with lower-is-better metrics inverted as (1–value), and the normalized scores were averaged. Uncertainty quantifies the prediction confidence of the trained random forest model, calculated as the standard deviation across individual tree predictions, where larger values indicate higher uncertainty. For reference, the average prediction uncertainty for compositions containing elements not included in the training dataset is 0.66, whereas it is only 0.28 for compositions whose elements are all present in the training data.

composition	cos θ	d_w	DEED (eV/atom) $^{-1}$	cost (USD/kg)	recycling rate (%)	abundance (mg/kg)	supply risk	source	uncertainty
Al–Cr–Cu–Fe–Zn	0.37	4.06	11.62	4.03	50.00	27 766.40	5.06	ML	0.69
Al–Cu–Fe–Sn–Zn	0.35	4.07	11.79	5.89	50.00	27 746.46	5.16	ML	0.70
Al–Cu–Fe–Ni–Zn	0.35	4.08	11.68	4.93	50.00	27 762.80	5.06	ML	0.76
Al–Co–Cu–Fe–Zn	0.36	4.04	11.83	8.71	50.00	27 751.00	5.34	ML	0.73
Ag–Al–Cu–Fe–Zn	0.34	4.18	11.69	106.35	50.00	27 746.02	5.06	ML	0.81
Al–Cr–Fe–Ni–Zn	0.36	4.15	11.57	5.61	50.00	27 771.20	5.44	ML	0.73
Al–Cr–Fe–Mn–Zn	0.36	4.22	11.58	3.20	50.00	27 944.40	5.34	ML	0.72
Al–Co–Fe–Ti–Zn	0.35	4.21	11.69	9.73	50.00	28 869.00	5.44	ML	0.75
Ag–Al–Fe–Sn–Zn	0.36	4.08	11.74	108.89	50.00	27 734.48	5.54	ML	0.71
Al–Fe–Ni–Sn–Zn	0.36	4.12	11.74	7.47	50.00	27 751.26	5.54	ML	0.72
Ag–Al–Fe–Ni–Zn	0.35	4.09	11.60	107.93	50.00	27 750.82	5.44	ML	0.75
Ag–Al–Co–Fe–Zn	0.36	4.05	11.79	111.71	50.00	27 739.02	5.72	ML	0.74
Al–Co–Fe–Ni–Zn	0.35	4.06	11.92	10.29	50.00	27 755.80	5.72	ML	0.71
Al–Co–Fe–Mn–Zn	0.35	4.14	11.60	7.88	50.00	27 929.00	5.62	ML	0.69
Al–Co–Cr–Fe–Zn	0.36	4.17	11.85	9.39	50.00	27 759.40	5.72	ML	0.72
Al–Co–Fe–Sn–Zn	0.36	4.14	11.74	11.25	50.00	27 739.46	5.82	ML	0.73
Cu–Fe–Ni–Ti–Zn	0.37	3.90	12.09	6.79	50.00	12 432.80	5.06	DFT	0.32
Al–Fe–Mo–Ni–Zn	0.36	4.15	11.57	11.75	46.00	27 751.04	5.92	ML	0.67
Al–Cr–Fe–Mo–Zn	0.36	4.25	11.86	10.85	46.00	27 754.64	5.92	ML	0.72
Al–Fe–Mo–Sn–Zn	0.36	4.17	11.51	12.71	46.00	27 734.70	6.02	ML	0.73
Al–Cu–Ni–Sn–Zn	0.35	4.06	11.96	8.59	50.00	16 503.26	5.36	ML	0.73
Ag–Al–Cu–Sn–Zn	0.35	4.07	11.95	110.01	50.00	16 486.48	5.36	ML	0.73
Al–Cr–Cu–Sn–Zn	0.37	4.14	11.52	7.69	50.00	16 506.86	5.36	ML	0.72
Co–Cu–Fe–Ti–Zn	0.37	3.85	11.98	10.57	50.00	12 421.00	5.34	DFT	0.46
Ag–Al–Cu–Ni–Zn	0.34	4.19	12.01	109.05	50.00	16 502.82	5.26	ML	0.80
Al–Co–Cr–Cu–Zn	0.37	4.06	11.48	10.51	50.00	16 511.40	5.54	ML	0.68
Al–Co–Cu–Sn–Zn	0.35	4.07	12.04	12.37	50.00	16 491.46	5.64	ML	0.72
Al–Co–Cu–Ni–Zn	0.34	4.17	12.10	11.41	50.00	16 507.80	5.54	ML	0.79
Ag–Al–Co–Cu–Zn	0.34	4.18	11.98	112.83	50.00	16 491.02	5.54	ML	0.79
Ag–Al–Ni–Sn–Zn	0.35	4.07	11.89	111.59	50.00	16 491.28	5.74	ML	0.72
Al–Co–Sn–Ti–Zn	0.35	4.24	11.63	13.39	50.00	17 609.46	5.74	ML	0.74
Al–Co–Cr–Ni–Zn	0.36	4.11	11.52	12.09	50.00	16 516.20	5.92	ML	0.73
Al–Co–Cr–Mn–Zn	0.36	4.20	11.60	9.67	50.00	16 689.40	5.82	ML	0.69
Ag–Al–Co–Sn–Zn	0.35	4.07	11.99	115.37	50.00	16 479.48	6.02	ML	0.72
Al–Co–Ni–Sn–Zn	0.35	4.11	11.94	13.95	50.00	16 496.26	6.02	ML	0.72
Al–Co–Cr–Sn–Zn	0.37	4.17	11.51	13.05	50.00	16 499.86	6.02	ML	0.66
Ag–Cu–Fe–Ti–Zn	0.35	4.48	11.64	108.21	50.00	12 416.02	5.06	ML	0.65
Ag–Al–Co–Ni–Zn	0.34	4.19	11.97	114.41	50.00	16 495.82	5.92	ML	0.78
Al–Cr–Mn–Mo–Zn	0.36	4.34	11.61	11.13	46.00	16 684.64	6.02	ML	0.65
Ag–Fe–Ni–Ti–Zn	0.34	4.63	11.50	109.79	50.00	12 420.81	5.44	ML	0.61

the reference data, indicating that the model operates within the same chemical domain. High cross-validation accuracy and consistent descriptor trends across all target properties further support the robustness of the model in the expanded space.

In applying the model, we also performed uncertainty quantification to assess the confidence of our predictions. As shown in table 1, predictions for compositions whose constituent elements are all represented in the training set exhibit lower uncertainties (average 0.28) than those including elements absent from the training data (average 0.66), where the model relies solely on ML. This increased uncertainty is expected, as the model must extrapolate beyond its original training domain when making predictions for compositions containing previously untrained elements. While ML enables rapid screening of candidate alloys, these results highlight the need for additional calculations and validation before recommending compositions with untrained elements for practical use. Importantly, our framework provides a systematic approach for identifying regions of greatest uncertainty, thereby offering a valuable path for

closed-loop experimental design; targeted experiments can be guided to reduce uncertainty, continually refining and improving predictive accuracy for novel alloy compositions.

Of the 40 top candidates identified via multi-objective optimization, 38 originated from the ML-predicted set, while only two came from the original DFT dataset. Candidate identification is based not only on ML-predicted values of $\cos\theta$, d_w , and DEED, but also on newly incorporated economic metrics. This enables simultaneous optimization across multiple objectives. While the key physicochemical property values of these candidates remain within the range observed in the training set, the addition of economic data and the multi-objective framework allow discovery of many compositions with superior overall catalyst suitability compared to those in the original seed data.

Future perspective. While this work demonstrates the potential of ML models for accurate prediction and evaluation of HEAs as fuel-cell catalysts, several key challenges remain.

First, ML predictions are typically confined to the distribution of the training data, as seen in figure 5, where property distributions for expanded compositional spaces mirror those from the seed dataset. To overcome this limitation and expand the model's applicability domain, future work should integrate active learning for targeted data augmentation [67–70]. In active learning, the model identifies the next most promising or uncertain compositions—those that would most improve model performance—for validation via new DFT calculations or experiments. This targeted approach prioritizes the acquisition of informative data. By iteratively retraining the model with these new samples, its ability to generalize improves, which in turn enables efficient exploration of complex or under-sampled regions.

Second, descriptors such as EFA and DEED have already demonstrated effectiveness in other chemistries, and their performance in metallic HEAs provides an exciting opportunity for further exploration and validation. Ongoing assessment—and, if needed, adaptation—of these or related descriptors, especially those tailored to the physics of HEA systems, could further improve predictive accuracy and deepen understanding of underlying mechanisms.

Further advances could also be achieved by incorporating higher-order computational methods to better capture the complex environments encountered in actual fuel-cell operation. For example, combining explicit solvation models [71, 72] and *ab-initio* molecular dynamics [68, 73, 74] can provide more accurate insights into the catalyst's behavior under realistic electrochemical conditions, including water and ion interactions at the catalyst/electrolyte interface [75, 76]. In addition, incorporating thermodynamic and kinetic modeling, such as calculating potential-dependent adsorption energies or simulating degradation pathways [77, 78], may offer a more complete understanding of catalytic activity, stability, and degradation mechanisms. These approaches, used in tandem with data-driven models, would enable more reliable predictions and guide the rational design of HEA catalysts optimized for operation in PEM fuel cells. Moreover, they would help inform how the proposed candidates meet practical requirements such as manufacturability, scalability, and industry standards.

A related research direction is phase stability and crystal structure preference—such as body-centered cubic versus face-centered cubic—evaluated using thermodynamic descriptors across a broad range of compositions. Although body-centered cubic is among the most prevalent parent structures for HEAs, many important examples, such as the prototypical CrMnFeCoNi alloy, crystallize in the face-centered cubic structure [79–81], and other HEAs form hexagonal close-packed or amorphous phases [60, 82, 83]. In this study, we restrict modeling to the body-centered cubic parent structure for consistency and clarity; however, our approach is directly applicable to other crystal structures. Future work will involve explicit application of the framework to additional parent structures, as well as benchmarking model predictions against experimentally confirmed single-phase HEAs to assess descriptor reliability and guide practical alloy discovery.

Another important direction is the integration of promising HEA catalyst compositions into practical PEM fuel-cell systems. Real-world implementation requires not only high activity and stability, but also compatibility with PEM operating environments, including resistance to corrosion, dissolution, and poisoning under acidic conditions [84, 85]. Future work should combine computational predictions with experimental testing in relevant device architectures, evaluating both intrinsic catalytic performance and long-term durability. In addition, understanding the interface between the HEA catalyst and PEM materials, as well as optimizing how the catalyst is incorporated into the membrane electrode assembly [5, 86], will be essential for effective translation of novel catalysts into functional fuel-cell devices.

Finally, the adoption of advanced MLIPs [87–92] and generative models [93–97] offers promising avenues for automated, AI-driven exploration and data generation. These techniques can facilitate the discovery of new regions within the high-entropy compositional space that remain inaccessible to conventional or brute-force approaches.

3. Conclusion

This work presents an effective, data-driven framework for rapid discovery of HEA catalysts for hydrogen fuel cells by integrating disorder-sensitive descriptors with ML to predict both synthesizability and catalytic performance. Most notably, we demonstrate that the DEED is a highly learnable metric for synthesizability in complex alloys, enabling accurate prediction from composition and proving more learnable than the EFA. In addition, this study introduces two novel DOS-based descriptors—cosine similarity ($\cos\theta$) and Wasserstein distance (d_w)—constructed to quantify catalytic activity in HEAs relative to platinum. These descriptors provide new tools for assessing electronic-structure similarity and screening for platinum-like catalytic performance.

Leveraging these innovations, we performed high-throughput screening of over 20 000 HEA compositions, identifying several platinum-free candidates that are economically viable, readily scalable, and exhibit promising predicted catalytic activity. These results demonstrate that disorder-sensitive descriptors—including both the DEED synthesizability metric and the DOS-based performance metrics—can be effectively incorporated into automated discovery pipelines, helping to overcome the configurational complexity of HEAs.

Our framework is broadly applicable to fields including energy conversion [28, 98], storage, manufacturing, and magnetics, where atomic-level disorder and supply-chain resilience are critical [13]. Automated screening using transferable, disorder-sensitive descriptors enables efficient identification of better performing and economically viable HEA candidates. By incorporating uncertainty quantification, our approach can highlight compositions and regions of greatest prediction uncertainty, guiding targeted experiments that maximize information gain. Integrating these methods into closed-loop workflows—where computational predictions and uncertainty metrics directly inform experimental validation—not only reduces time to discovery and deployment of new materials [99], but also iteratively improves model accuracy through continual feedback. Continued refinement of disorder-sensitive metrics and deeper integration with uncertainty-guided experimental workflows will further strengthen this approach and expand its impact within and beyond alloy systems.

4. Methods

HEA modeling. HEAs were treated within the partial-occupation module (POCC) of the automatic flow Framework for materials discovery (aflow++) [29, 30]. For each equiatomic A–B–C–D–E composition, the module generates 600 five-atom supercell tiles based on the body-centered cubic parent structure (bcc; spacegroup #229), 49 of which it determines to be unique by symmetry [29, 30, 100]. The total energy and derived properties of the unique tiles were obtained from DFT calculations performed with the Vienna Ab initio Simulation Package (VASP) [64], employing projector-augmented wave (PAW) pseudopotentials [101, 102] and the Perdew–Burke–Ernzerhof (PBE) generalized-gradient approximation (GGA) [103]. All additional parameters and convergence criteria are as detailed in [104]. Based on the $n=49$ DFT calculations, the EFA of the composition is computed as:

$$\text{EFA} = \sigma \{H_i\}^{-1} = \left(\sqrt{\frac{\sum_{i=1}^n g_i (H_i - H_{\text{mix}})^2}{(\sum_{i=1}^n g_i) - 1}} \right)^{-1}, \quad (1)$$

where H_i is the i -tile's enthalpy (from DFT), g_i is the i -tile's degeneracy (from POCC's structural-uniqueness analysis), and H_{mix} is the mean enthalpy: $H_{\text{mix}} = (\sum_{i=1}^n g_i H_i) / (\sum_{i=1}^n g_i)$ [16]. 48 of the 49 structures were determined to have a degeneracy of 10, and the other a degeneracy of 120. The properties—including EFA, g_i 's, H_{mix} —are computed automatically by afloop-POCC's postprocessing procedure and written to the `afloop.pocc.out` file. Example commands for generating the input files and performing the postprocessing for this analysis automatically are provided in section 8 ‘Modeling disorder’ of [29].

A total of 427 HEA compositions, drawn from combinations of 11 commonly studied transition metals (Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Mo, and W) [83, 105], were fully characterized using afloop-POCC and DFT. As a proof-of-concept, only the bcc parent structure was considered in this work. Since HEAs commonly crystallize in high-symmetry structures—bcc being among the most prevalent [12, 15, 83, 105–107]—the scope facilitates an initial investigation of the relationship between composition and key thermodynamic descriptors, such as EFA, distance to the convex hull, and DEED. Future work, especially in the context of experimental validation, should include competing high-entropy phases based on alternative parent structures. Such extension is feasible within the current supercell-ensemble framework, with additional calculations.

The supercell ensemble approach implemented in `aflow`-POCC offers a distinct alternative to the widely used special quasirandom structure method (SQS) [108]. SQS aims to model the disordered system using a single, carefully constructed supercell that minimizes pair and higher-order site correlations, providing an effective approximation to infinite-temperature disorder. This typically results in the use of large, low-symmetry supercells. In contrast, `aflow`-POCC represents disorder using multiple small-cell supercells, with ensemble averages calculated via Boltzmann weighting, thereby explicitly considering energy, temperature, and structural degeneracy. To avoid redundant configurations, each candidate supercell is evaluated using the universal Force Field [100]. Degeneracy is determined by comparing the resulting energies, which allows for efficient identification of symmetry-distinct structures without exhaustive enumeration [30]. The small size of the supercells also makes high-throughput calculations more feasible, enabling efficient parallelization across computing resources. Notably, ensemble averaging with `aflow`-POCC has demonstrated particular effectiveness at finite temperatures [31, 50, 109, 110]. While some SQS-based workflows attempt to capture configurational effects by averaging over multiple SQSs, the core SQS objective function [111] remains oriented toward the high-temperature limit. By comparison, `aflow`-POCC provides a tunable synthesis temperature parameter to sample the relevant thermodynamic regime for a given application. For a more detailed discussion and further methodological comparisons, see Section 8 ‘Modeling disorder’ subsection ‘Comparison to other disordered system models’ of [29].

Convex hull analysis. To quantify thermodynamic synthesizability, we calculated DEED = $\sqrt{\text{EFA}/\langle \Delta H_{\text{hull}} \rangle}$, which requires the HEA’s ensemble-average distance to the convex hull ($\langle \Delta H_{\text{hull}} \rangle$). This quantity measures the enthalpic penalty for forming the high-entropy phase relative to the most stable set of competing ordered phases. Quinary convex hulls and related properties were computed using the convex hull module in `aflow++` (`aflow`-CHULL) [34]. Accurate *ab-initio* convex hull construction depends on a comprehensive database of energetically competing phases in each chemical system.

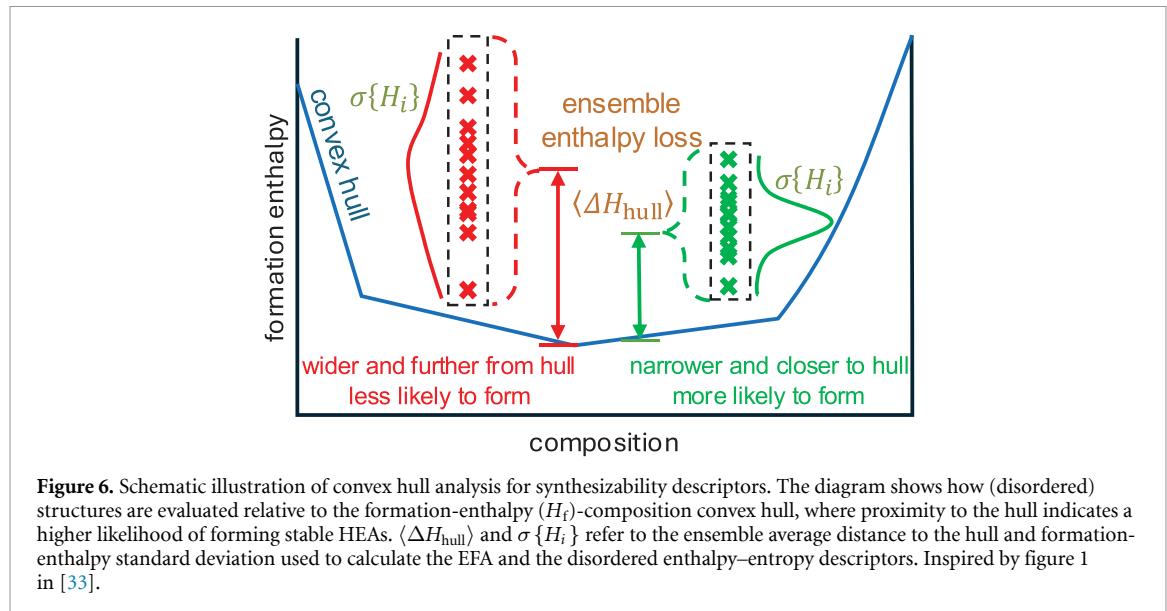
Data for each hull were assembled using two strategies. First, we included known ordered binary and ternary compounds from the inorganic crystal structure database (ICSD) [112], recalculated with standardized DFT protocols [104]. The most critical contributions to the enthalpic penalty come from the binary and ternary compounds, with decreasing importance for more complex chemistries. As supported by previous studies [17, 113], higher-order (quaternary, quinary) phases contribute negligibly to the enthalpy landscape for HEA systems on average. This is because stable binary and ternary phases act as strong competitors, requiring any higher-order phase to be more stable than all relevant lower-order phases. Consequently, despite the vast number of possible higher-order compositions—growing combinatorially with the number of elements—only very few of them are actually stable. Second, we expanded phase coverage by redecorating common crystallographic prototypes for metallic systems [114–116] with combinations of the 11 transition metals in this study, generating a large set of hypothetical structures.

The resulting DFT energies, from both experimental and hypothetical structures, formed the basis for the convex hull phase diagrams. Each quinary hull contained over 2000 structures, ensuring convergence of the thermodynamic properties. The `aflow`-CHULL module identified the stable ground-state phases that define the minimum energy tie-surfaces (the convex hull). From this, we calculated $\langle \Delta H_{\text{hull}} \rangle$ as the energetic distance between the convex hull and the centroid formation enthalpy of the supercell ensemble representing the HEA [33, 34]. A schematic illustration of this concept is provided in figure 6, where the relative position of compositions with respect to the formation-energy hull serves as the basis for our synthesizability descriptor. This calculation is facilitated by the `--hull_enthalpy` flag as described in Section 6 ‘Thermodynamics’ of [29].

DOS analysis. To model the electronic structure of disordered systems, we applied a statistical ensemble approach using supercell tiles generated by the `aflow`-POCC framework. The total DOS and orbital-projections, including the *d*-DOS, were calculated for each unique supercell tile using DFT. The DOS and orbital-projections of the disordered HEA were then obtained as ensemble averages over all tiles, with each tile weighted by its Boltzmann probability and degeneracy. The Boltzmann probability for the *i*-tile is given by:

$$P_i = \frac{g_i \exp(- (H_i - H_{\text{gs}}) / k_B T)}{\sum_{i=1}^n g_i \exp(- (H_i - H_{\text{gs}}) / k_B T)}, \quad (2)$$

where H_{gs} is the minimum enthalpy of the supercell ensemble and T is the HEA synthesis temperature [29, 30]. Explicitly, the disordered *d*-DOS is calculated as an ensemble average over the



d -DOS of the tiles: $d\text{-DOS}(E_j) = \sum_{i=1}^n P_i d\text{-DOS}_i(E_j)$, where E_j denotes the j -energy value on the grid used for calculating the DOS.

The `aflow`-POCC framework employs a Boltzmann statistics approach to ensemble-average the properties of supercell configurations to describe disordered systems. At low temperatures, the ground-state structures dominate, whereas at high temperatures the system approaches the annealing limit where all configurations are considered equiprobable. We evaluated the relative change of the electronic-structure descriptors with respect to 1500 K across different synthesis temperatures. As shown in figure 7, for $T > 1200$ K the variations in $\cos\theta$ and d_w shrink below $\sim 0.5\%$ and become effectively negligible, indicating that the system is equilibrated in this regime with respect to this approach. Accordingly, we select 1500 K as the reference point as it is a typical synthesis temperature for many HEAs and remains below reported melting temperatures [107, 117–119]. The synthesis temperature is distinct from the operational temperature, and the descriptors in this regime are only weakly sensitive to temperature, minimizing thermal bias in the ML model. We note that the present treatment captures configurational disorder and qualitative trends; more accurate results at elevated temperatures could be obtained by incorporating vibrational free-energy contributions, which can be achieved within the `aflow`-POCC framework with additional calculations [31, 109].

To assess the catalytic activity of HEAs, we compare their ensemble-averaged d -DOS profiles with that of Pt, a well-established benchmark catalyst. The DOS of platinum was resolved using DFT and shows good agreement with published theoretical calculations and photoemission measurements [120–122]. The similarity between DOS profiles is quantified using the cosine similarity metric ($\cos\theta$), defined as:

$$\cos\theta = \frac{\int_{-\infty}^{\infty} d\text{-DOS}_{\text{cand}}(E) d\text{-DOS}_{\text{Pt}}(E) dE}{\sqrt{\int_{-\infty}^{\infty} d\text{-DOS}_{\text{cand}}^2(E) dE} \sqrt{\int_{-\infty}^{\infty} d\text{-DOS}_{\text{Pt}}^2(E) dE}}, \quad (3)$$

where $d\text{-DOS}_{\text{cand}}(E)$ and $d\text{-DOS}_{\text{Pt}}(E)$ denote the normalized d -DOS of the candidate alloy and platinum, respectively. Larger values of $\cos\theta$ indicate greater electronic resemblance to Pt, suggesting favorable catalytic properties.

In addition to $\cos\theta$, we also compute the Wasserstein distance (d_w) between the d -DOS of each candidate and that of Pt as a measure of dissimilarity. The Wasserstein distance is the minimum total ‘transport cost’—the integral of probability density moved multiplied by the distance it travels—required to transform one DOS profile into another. A smaller d_w value corresponds to a closer match to Pt, complementing $\cos\theta$ by taking into account both the overall shape and energy alignment of the DOS. The Wasserstein distance is calculated as:

$$d_w = \int_{-\infty}^{\infty} |\text{DOS}_{\text{cand}}^{\text{CDF}}(E) - \text{DOS}_{\text{Pt}}^{\text{CDF}}(E)| dE, \quad (4)$$

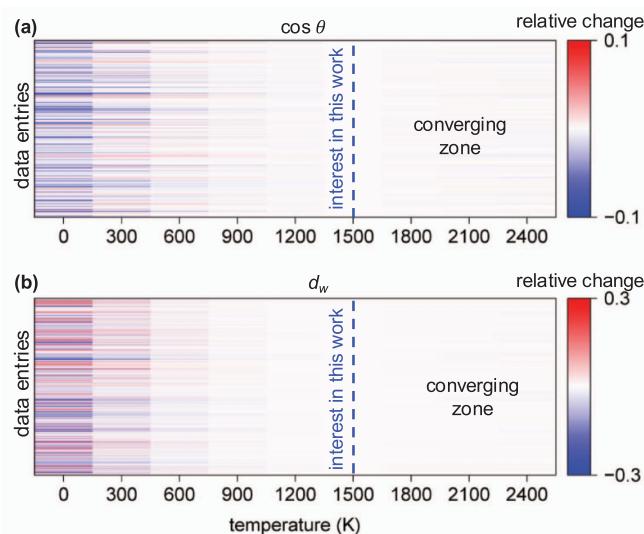


Figure 7. Sensitivity of electronic-structure descriptors to synthesis temperature. Relative changes of the electronic structure descriptors with respect to 1500 K are shown.

where $DOS_{\text{cand}}^{\text{CDF}}(E)$ and $DOS_{\text{Pt}}^{\text{CDF}}(E)$ denote the cumulative distribution functions (CDF) of the d -DOS for the candidate alloy and Pt, respectively.

ML modeling. We developed supervised ML models to predict four DFT-based descriptors that gauge the synthesizability and catalytic performance of HEAs: EFA, DEED, $\cos\theta$, and d_w . Elemental properties were extracted from the `aflow-xElement` database [29], including atomic number, atomic mass, electronegativity, atomic radii, melting point, boiling temperature, and atomization enthalpy. For each alloy, statistical descriptors—including minimum, maximum, average, max-to-min ratio, and root-mean-square deviation (atomic mismatch)—were calculated across the constituent elements to generate the input feature set. Feature construction was guided by descriptors commonly used for HEAs [37, 123, 124]. Random-forest regression models were trained with these features as inputs and the DFT-derived EFA, DEED, $\cos\theta$, and d_w values as outputs. Model performance was quantified by the MAE for accuracy and the coefficient of determination (r^2) for ranking capability, enabling rapid screening and prioritization of promising HEA compositions.

ML-driven screening. We selected 22 commonly used metallic elements as HEA candidates and generated all combinatorial chemical compositions [60, 61, 83, 105]. Using the trained machine-learning models, we predicted key material descriptors for synthesizability and catalytic performance across 26 334 unique alloy compositions. These predictions were then integrated with economic data—including crustal abundance, elemental price, recycling rate (discretized as 10 for <10%, 30 for 10%–30%, and 50 for >30%), and supply risk—extracted from the `mendeleev` Python library [63]. Missing values were imputed using the median of the dataset. This approach enabled construction of a comprehensive, multidimensional database to support accelerated discovery and informed prioritization of promising HEA candidates with balanced properties, economic viability, and scalable manufacturing considerations.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/entropy4energy/2025_HE_FuelCell_Discovery.

Acknowledgment

The authors thank Yushan Yan (University of Delaware), Bill McNamara (JHU), A J Donelson (JHU), Chip Cotton (JHU), Matthew Brownrigg (JHU), Katie Cariaga (JHU), Maya Niyogi (JHU), Will Shiber (JHU), and Shao-Yu Tseng (JHU) for helpful discussions. Research was sponsored by the Advanced Research Projects Agency-Energy (ARPA-E) (DE-AR0001787). X X and A A acknowledge support from the Institute for Data-Intensive Engineering and Science (IDIES) at JHU and the Space@Hopkins seed Grant fund. J L acknowledges support from the Johns Hopkins Data Science and AI Institute (DSAI). G Q acknowledges support from the Ralph O'Connor Sustainable Energy Institute (ROSEI) at JHU. C O

acknowledges the computational resources provided by the Advanced Research Computing at Hopkins (ARCH), which is supported by the National Science Foundation (NSF) Grant number OAC1920103, and the Texas Advanced Computing Center (TACC) at The University of Texas at Austin.

Conflicts of interest

The authors declare no competing financial interest.

Author contributions

Guangshuai Han  0000-0001-9297-5072

Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Software (equal), Validation (equal), Visualization (equal), Writing – original draft (equal), Writing – review & editing (equal)

Tianhao Li  0009-0006-2035-7754

Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Software (equal), Validation (equal), Visualization (equal), Writing – original draft (equal), Writing – review & editing (equal)

Xiao Xu  0009-0008-4118-4746

Formal analysis (supporting), Investigation (supporting), Methodology (supporting), Software (supporting), Validation (supporting), Visualization (supporting), Writing – original draft (supporting), Writing – review & editing (supporting)

Jaehyung Lee  0009-0005-4280-9144

Formal analysis (supporting), Investigation (supporting), Methodology (supporting), Software (supporting), Validation (supporting), Visualization (supporting), Writing – original draft (supporting), Writing – review & editing (supporting)

Guotao Qiu

Formal analysis (supporting), Investigation (supporting), Methodology (supporting), Software (supporting), Validation (supporting), Visualization (supporting), Writing – original draft (supporting), Writing – review & editing (supporting)

Sabrina Sequeira  0009-0001-1456-9711

Formal analysis (supporting), Investigation (supporting), Methodology (supporting), Software (supporting), Validation (supporting), Visualization (supporting), Writing – original draft (supporting), Writing – review & editing (supporting)

Akshaya Ajith  0009-0008-5214-183X

Data curation (supporting), Formal analysis (supporting), Methodology (supporting), Software (supporting)

Corey Oses  0000-0002-3790-1377

Conceptualization (equal), Data curation (equal), Funding acquisition (equal), Methodology (equal), Project administration (equal), Resources (equal), Software (equal), Supervision (equal), Writing – original draft (equal), Writing – review & editing (equal)

References

- [1] Alaswad A, Omran A, Sodre J R, Wilberforce T, Pignatelli G, Dassisti M, Baroutaji A and Olabi A G 2021 Technical and commercial challenges of proton-exchange membrane (PEM) fuel cells *Energies* **14** 144
- [2] Chen Z, Waje M, Li W and Yan Y 2007 Supportless Pt and PtPd nanotubes as electrocatalysts for oxygen-reduction reactions *Angew. Chem.* **46** 4060–3
- [3] Cao G, Yang S, Ren J-C and Liu W 2025 Electronic descriptors for designing high-entropy alloy electrocatalysts by leveraging local chemical environments *Nat. Commun.* **16** 1251
- [4] Liu M, Zhao Z, Duan X and Huang Y 2019 Nanoscale structure design for high-performance Pt-based ORR catalysts *Adv. Mater.* **31** 1802234
- [5] Gao F-Y et al 2022 Nickel–molybdenum–niobium metallic glass for efficient hydrogen oxidation in hydroxide exchange membrane fuel cells *Nat. Catal.* **5** 993–1005
- [6] Zitolo A, Goellner V, Armel V, Sougrati M-T, Mineva T, Stievano L, Fonda E and Jaouen F 2015 Identification of catalytic sites for oxygen reduction in iron- and nitrogen-doped graphene materials *Nat. Mater.* **14** 937–42
- [7] Kabe Y et al 2016 Haem-dependent dimerization of PGRMC1/Sigma-2 receptor facilitates cancer proliferation and chemoresistance *Nat. Commun.* **7** 11030

- [8] Cheng C-F, Li X, Liu K, Zou F, Tung W-Y, Huang Y-F, Xia X, Wang C-L, Vogt B D and Zhu Y 2019 A high-performance lithium-ion capacitor with carbonized NiCo_2O_4 anode and vertically-aligned carbon nanoflakes cathode *Energy Storage Mater.* **22** 265–74
- [9] Krishnan S R, Verstraete D and Aguey-Zinsou F 2024 Performance of non-precious metal electrocatalysts in proton-exchange membrane fuel cells: a review *ChemElectroChem* **11** e202400299
- [10] Xu X, Kang N, Zulevi B, Serov A and Pintauro P N 2023 Highly durable platinum group metal-free catalyst fiber cathode MEAs for proton exchange membrane fuel cells *J. Power Sources* **586** 233679
- [11] Osmieri L and Meyer Q 2022 Recent advances in integrating platinum group metal-free catalysts in proton exchange membrane fuel cells *Curr. Opin. Electrochem.* **31** 100847
- [12] Oses C, Toher C and Curtarolo S 2020 High-entropy ceramics *Nat. Rev. Mater.* **5** 295–309
- [13] Qiu G, Li T, Xu X, Liu Y, Niyogi M, Cariaga K and Oses C 2025 High entropy powering green energy: hydrogen, batteries, electronics and catalysis *npj Comput. Mater.* **11** 145
- [14] Yeh J-W, Chen S-K, Lin S-J, Gan J-Y, Chin T-S, Shun T-T, Tsau C-H and Chang S-Y 2004 Nanostructured high-entropy alloys with multiple principle elements: novel alloy design concepts and outcomes *Adv. Eng. Mater.* **6** 299–303
- [15] George E P, Raabe D and Ritchie R O 2019 High-entropy alloys *Nat. Rev. Mater.* **4** 515–34
- [16] Sarker P, Harrington T, Toher C, Oses C, Samee M, Maria J-P, Brenner D W, Vecchio K S and Curtarolo S 2018 High-entropy high-hardness metal carbides discovered by entropy descriptors *Nat. Commun.* **9** 4980
- [17] Toher C, Oses C, Hicks D and Curtarolo S 2019 Unavoidable disorder and entropy in multi-component systems *npj Comput. Mater.* **5** 69
- [18] Rao Z et al 2022 Machine learning-enabled high-entropy alloy discovery *Science* **378** 78–85
- [19] Vela B, Khatamsaz D, Acemi C, Karaman I and Arróyave R 2023 Data-augmented modeling for yield strength of refractory high entropy alloys: a bayesian approach *Acta Mater.* **261** 119351
- [20] Feng R et al 2021 High-throughput design of high-performance lightweight high-entropy alloys *Nat. Commun.* **12** 4329
- [21] Yan Y, Hu X, Liao Y, Zhou Y, He W and Zhou T 2025 Recent machine learning-driven investigations into high entropy alloys: a comprehensive review *J. Alloys Compd.* **1010** 177823
- [22] Chen K and Min B-W 2025 Predicting hardness in high entropy alloys with explainable machine learning *Mater. Today Commun.* **45** 112388
- [23] Liu Y, Yao Z, Zhang P, Xu Z, Lin S, He M, Lu S and Wu X 2022 Tailoring high-temperature oxidation resistance of FeCrMnVSix high entropy alloy coatings via building Si-rich dendrite microstructure *Appl. Surf. Sci.* **606** 154862
- [24] Mishin Y 2021 Machine-learning interatomic potentials for materials science *Acta Mater.* **214** 116980
- [25] Ram S, Lee A S, Lee S-C and Bhattacharjee S 2025 Advanced multifunctional electrocatalysts: integrating DFT and machine learning for OER, HER and ORR reactions *Chem. Mater.* **37** 3608–21
- [26] Chen Z-J et al 2021 Engineering microdomains of oxides in high-entropy alloy electrodes toward efficient oxygen evolution *Adv. Mater.* **33** 2101845
- [27] Yoon U, Jeong K and Kim S H 2025 Advancing electrocatalysis through density functional theory: From reaction mechanisms to machine learning integration *J. CO₂ Util.* **101** 103192
- [28] Oses C, Li T, Xu X, Han G, Qiu G and Owens J R 2025 Beyond the four core effects: revisiting thermoelectrics with a high-entropy design *Mater. Horiz.* **12** 5946–56
- [29] Oses C et al 2023 aflow++: a C++ framework for autonomous materials design *Comput. Mater. Sci.* **217** 111889
- [30] Yang K, Oses C and Curtarolo S 2016 Modeling off-stoichiometry materials with a high-throughput *Ab-Initio* approach *Chem. Mater.* **28** 6484–92
- [31] Esters M, Smolyanyuk A, Oses C, Hicks D, Divilov S, Eckert H, Campilongo X, Toher C and Curtarolo S 2023 QH-POCC: taming tiling entropy in thermal expansion calculations of disordered materials *Acta Mater.* **245** 118594
- [32] Liu Z K 2024 *Zentropy: Theory and Fundamentals* (Jenny Stanford Publishing)
- [33] Divilov S et al 2024 Disordered enthalpy-entropy descriptor for high-entropy ceramics discovery *Nature* **625** 66–73
- [34] Oses C et al 2018 AFLOW-CHULL: cloud-oriented platform for autonomous phase stability analysis *J. Chem. Inf. Model.* **58** 2477–90
- [35] Perim E et al 2016 Spectral descriptors for bulk metallic glasses based on the thermodynamics of competing crystalline phases *Nat. Commun.* **7** 12315
- [36] Guo S and Liu C T 2011 Phase stability in high entropy alloys: formation of solid-solution phase or amorphous phase *Prog. Nat. Sci.-Mater.* **21** 433–46
- [37] Shannon R D 1976 Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides *Acta Crystallogr. Sect. A* **32** 751–67
- [38] Callister Jr W D 2006 *Materials Science and Engineering: an Introduction* 7th edn (Wiley)
- [39] Nørskov J K, Abild-Pedersen F, Studt F and Bligaard T 2011 Density functional theory in surface chemistry and catalysis *Proc. Natl Acad. Sci.* **108** 937–43
- [40] Hammer B and Nørskov J K 2000 Theoretical surface science and catalysis—calculations and concepts *Adv. Catal.* **45** 71–129
- [41] Jiao S, Fu X and Huang H 2022 Descriptors for the evaluation of electrocatalytic reactions: d-band theory and beyond *Adv. Funct. Mater.* **32** 2107651
- [42] Xu S, Jiao D, Ruan X, Jin Z, Qiu Y, Fan J, Zhang L, Zheng W and Cui X 2024 Synergistic modulation of the d-band center in Ni_3S_2 by selenium and iron for enhanced oxygen evolution reaction (OER) and urea oxidation reaction (UOR) *J. Colloid Interface Sci.* **671** 46–55
- [43] Wang Z, Shen S, Wang J and Zhong W 2024 Modulating the D-Band Center of Electrocatalysts for Enhanced Water Splitting *Chem. Eur. J.* **30** e202402725
- [44] Bhattacharjee S, Waghmare U V and Lee S-C 2016 An improved d-band model of the catalytic activity of magnetic transition metal surfaces *Sci. Rep.* **6** 35916
- [45] Darby M T, Stamatakis M, Michaelides A and Sykes E C H 2018 Lonely atoms with special gifts: breaking linear scaling relationships in heterogeneous catalysis with single-atom alloys *J. Phys. Chem. Lett.* **9** 5636–46
- [46] Goda A M, Barteau M A and Chen J G 2006 Correlating electronic properties of bimetallic surfaces with reaction pathways of C_2 hydrocarbons *J. Phys. Chem. B* **110** 11823–31
- [47] Calle-Vallejo F, Koper M T M and Bandarenka A S 2013 Tailoring the catalytic activity of electrodes with monolayer amounts of foreign metals *Chem. Soc. Rev.* **42** 5210–30

- [48] Han J, Kamber M and Pei J 2012 2 - Getting to Know Your Data *Data Mining: Concepts and Techniques* the Morgan Kaufmann Series in Data Management Systems 3rd edn, ed J Han, M Kamber and J Pei (Morgan Kaufmann) pp 39–82
- [49] Shen J, Qu Y, Zhang W and Yu Y 2018 Wasserstein distance guided representation learning for domain adaptation *Proc. AAAI conf. on artificial Intelligence* vol 32
- [50] Hossain M D et al 2021 Entropy landscaping of high-entropy carbides *Adv. Mater.* **33** 2102904
- [51] Hossain M D et al 2021 Carbon stoichiometry and mechanical properties of high entropy carbides *Acta Mater.* **215** 117051
- [52] Vecchio K, Curtarolo S, Kaufmann K, Harrington T J, Oses C and Toher 2024 Fermi energy engineering of enhanced plasticity in high-entropy carbides *Acta Mater.* **276** 120117
- [53] Sun Y and Dai S 2021 High-entropy materials for catalysis: a new frontier *Sci. Adv.* **7** eabg1600
- [54] Deng C, Wang T, Wu P, Zhu W and Dai S 2023 High entropy materials for catalysis: a critical review of fundamental concepts and applications *Nano Energy* **120** 109153
- [55] Pedregosa F et al 2011 Scikit-learn: machine learning in python *J. Mach. Learn. Res.* **12** 2825–30
- [56] Isayev O, Fourches D, Muratov E N, Oses C, Rasch K, Tropsha A and Curtarolo S 2015 Materials cartography: representing and mining materials space using structural and electronic fingerprints *Chem. Mater.* **27** 735–43
- [57] Stanev V, Oses C, Kusne A G, Rodriguez E, Paglione J, Curtarolo S and Takeuchi I 2018 Machine learning modeling of superconducting critical temperature *npj Comput. Mater.* **4** 29
- [58] Kaufmann K, Maryanovsky D, Mellor W M, Zhu C, Rosengarten A S, Harrington T J, Oses C, Toher C, Curtarolo S and Vecchio K S 2020 Discovery of high-entropy ceramics via machine learning *npj Comput. Mater.* **6** 42
- [59] Kaufmann K and Vecchio K S 2020 Searching for high entropy alloys: a machine learning approach *Acta Mater.* **198** 178–222
- [60] Miracle D B and Senkov O N 2017 A critical review of high entropy alloys and related concepts *Acta Mater.* **122** 448–511
- [61] Ye Y F, Wang Q, Lu J, Liu C T and Yang Y 2016 High-entropy alloy: challenges and prospects *Mater. Today* **19** 349–62
- [62] Han L, Mu W, Wei S, Liaw P K and Raabe D 2024 Sustainable high-entropy materials? *Sci. Adv.* **10** eads3926
- [63] M T Hacking 2024 Mendeleev: a python resource for properties of chemical elements (available at: <https://mendeleev.readthedocs.io/en/stable/>) (Accessed 24 June 2025)
- [64] Kresse G and Furthmüller J 1996 Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set *Phys. Rev. B* **54** 11169–86
- [65] Parr R G and Weitao Y 1994 *Density-Functional Theory of Atoms and Molecules* (International Series of Monographs on Chemistry) (Oxford University Press)
- [66] Kaufman L and Bernstein H 1970 *Computer Calculation of Phase Diagrams With Special Reference to Refractory Metals* (Refractory Materials Academic)
- [67] Isayev O, Oses C, Toher C, Gossett E, Curtarolo S and Tropsha A 2017 Universal fragment descriptors for predicting properties of inorganic crystals *Nat. Commun.* **8** 15679
- [68] Wang C, Aoyagi K, Wisesa P and Mueller T 2020 Lithium ion conduction in cathode coating materials from on-the-fly machine learning *Chem. Mater.* **32** 3741–52
- [69] Leverant C J and Harvey J A 2024 Accelerating the discovery of new, single phase high entropy ceramics via active learning *Chem. Mater.* **36** 10994–1003
- [70] Nie S, Xiang Y, Wu L, Lin G, Liu Q, Chu S and Wang X 2024 Active learning guided discovery of high entropy oxides featuring high H₂-production *J. Am. Chem. Soc.* **146** 29325–34
- [71] Ringe S, Hörmann N G, Oberhofer H and Reuter K 2022 Implicit solvation methods for catalysis at electrified interfaces *Chem. Rev.* **122** 10777–820
- [72] Jinnouchi R, Karsai F and Kresse G 2025 Absolute standard hydrogen electrode potential and redox potentials of atoms and molecules: machine learning aided first principles calculations *Chem. Sci.* **16** 2335–43
- [73] Yang G et al 2025 Theoretical insights on transition metals confined in black phosphorus for engineering stable single-atom catalysts *Surf. Interfaces* **70** 106825
- [74] Gardini A T, Raucci U and Parrinello M 2025 Machine learning-driven molecular dynamics unveils a bulk phase transformation driving ammonia synthesis on barium hydride *Nat. Commun.* **16** 2475
- [75] Zhang R, Pearce P E, Duan Y, Dubouis N, Marchandier T and Grimaud A 2019 Importance of water structure and catalyst-electrolyte interface on the design of water splitting catalysts *Chem. Mater.* **31** 8248–59
- [76] Khani H, Santiago A R P and He T 2023 An interfacial view of cation effects on electrocatalysis systems *Angew. Chem. Int. Ed.* **62** e202306103
- [77] Zhang Z, Li J and Wang Y-G 2024 Modeling interfacial dynamics on single atom electrocatalysts: explicit solvation and potential dependence *Acc. Chem. Res.* **57** 198–207
- [78] Li B, Holby E F and Wang G 2022 Mechanistic insights into metal, nitrogen doped carbon catalysts for oxygen reduction: progress in computational modeling *J. Mater. Chem. A* **10** 23959–72
- [79] Cantor B, Chang I T H, Knight P and Vincent A J B 2004 Microstructural development in equiatomic multicomponent alloys *Mater. Sci. Eng. A* **375-377** 213–8
- [80] Zhao S, Li Z, Zhu C, Yang W, Zhang Z, Armstrong D E J, Grant P S, Ritchie R O and Meyers M A 2021 Amorphization in extreme deformation of the CrMnFeCoNi high-entropy alloy *Sci. Adv.* **7** eabb3108
- [81] Gromov V E, Konovalov S V, Ivanov Y F, Shliarova Y A, Vorobyov S V and Semin A P 2022 Structure and properties of the CrMnFeCoNi high-entropy alloy irradiated with a pulsed electron beam *J. Mater. Res. Technol.* **19** 4258–69
- [82] Miracle D B 2017 High-entropy alloys: a current evaluation of founding ideas and core effects and exploring nonlinear alloys *JOM* **69** 2130–6
- [83] Senkov O N, Miracle D B, Chaput K J and Couzinie J-P 2018 Development and exploration of refractory high entropy alloys—a review *J. Mater. Res.* **33** 3092–128
- [84] Gu S, Cai R and Yan Y 2011 Self-crosslinking for dimensionally stable and solvent-resistant quaternary phosphonium based hydroxide exchange membranes *Chem. Commun.* **47** 2856–8
- [85] Zhang F, Zu B, Wang B, Qin Z, Yao J, Wang Z, Fan L and Jiao K 2025 Developing long-durability proton-exchange membrane fuel cells *Joule* **9** 101853
- [86] Sheng W, Bivens A P, Myint M N Z, Zhuang Z, Forest R V, Fang Q, Chen J G and Yan Y 2014 Non-precious metal electrocatalysts with high activity for hydrogen oxidation reaction in alkaline electrolytes *Energy Environ. Sci.* **7** 1719–24
- [87] Butler K T, Davies D W, Cartwright H, Isayev O and Walsh A 2018 Machine learning for molecular and materials science *Nature* **559** 547–55

- [88] Deng B, Zhong P, Jun K, Riebesell J, Han K, Bartel C J and Ceder G 2023 CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling *Nat. Mach. Intell.* **5** 1031–41
- [89] Zhu S *et al* 2025 Machine learning potentials for alloys: a detailed workflow to predict phase diagrams and benchmark accuracy (arXiv:2506.16771)
- [90] Wang H *et al* 2018 DeePMD-kit: a deep learning package for many-body potential energy representation and molecular dynamics *Comput. Phys. Commun.* **228** 178–84
- [91] Batzner S, Musaelian A, Sun L, Geiger M, Mailoa J P, Kornbluth M, Molinari N, Smidt T E and Kozinsky B 2022 E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials *Nat. Commun.* **13** 2453
- [92] Willman J T, Nguyen-Cong K, Williams A S, Belonoshko A B, Moore S G, Thompson A P, Wood M A and Oleynik I I 2022 Machine learning interatomic potential for simulations of carbon at extreme conditions *Phys. Rev. B* **106** L180101
- [93] Zeni C *et al* 2025 A generative model for inorganic materials design *Nature* **639** 624–32
- [94] Li H *et al* 2025 Conditional generative modeling for amorphous multi-element materials (arXiv:2503.07043)
- [95] Zhao Y, Al-Fahdi M, Hu M, Siriwardane E M D, Song Y, Nasiri A and Hu J 2021 High-throughput discovery of novel cubic crystal materials using deep generative neural networks *Adv. Sci.* **8** 2100566
- [96] Wang Z and You F 2025 Leveraging generative models with periodicity-aware, invertible and invariant representations for crystalline materials design *Nat. Comput. Sci.* **5** 365–76
- [97] Chen Z, Yuan Y, Zheng S, Guo J, Liang S, Wang Y and Wang Z 2025 Transformer-enhanced variational autoencoder for crystal structure prediction (arXiv:2502.09423)
- [98] Gong T, Qiu G, He M-R, Safanova O V, Yang W-C, Raciti D, Oses C and Hall A S 2024 Atomic ordering-induced ensemble variation in alloys governs electrocatalyst on/off states *J. Am. Chem. Soc.* **147** 510–8
- [99] Kusne A G *et al* 2020 On-the-fly closed-loop materials discovery via Bayesian active learning *Nat. Commun.* **11** 5966
- [100] Rappe A K, Casewit C J, Colwell K S and Goddard III W A 1992 UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations *J. Am. Chem. Soc.* **114** 10024–35
- [101] Kresse G and Furthmüller J 1996 Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set *Comput. Mater. Sci.* **6** 15–50
- [102] Kresse G and Joubert D 1999 From ultrasoft pseudopotentials to the projector augmented-wave method *Phys. Rev. B* **59** 1758–75
- [103] Perdew J P, Burke K and Ernzerhof M 1996 Generalized Gradient Approximation Made Simple *Phys. Rev. Lett.* **77** 3865–8
- [104] Calderon C E *et al* 2015 The AFLOW standard for high-throughput materials science calculations *Comput. Mater. Sci. A* **108** 233–8
- [105] Senkov O N, Miller J D, Miracle D B and Woodward C 2015 Accelerated exploration of multi-principal element alloys with solid solution phases *Nat. Commun.* **6** 6529
- [106] Jankousky M, Chen H, Novick A and Stevanović V 2024 All ‘roads’ lead to rocksalt structure (arXiv:2402.06843)
- [107] Chen W, Hilhorst A, Bokas G, Gorsse S, Jacques P J and Hautier G 2023 A map of single-phase high-entropy alloys *Nat. Commun.* **14** 2856
- [108] Zunger A, Wei S-H, Ferreira L G and Bernard J E 1990 Special quasirandom structures *Phys. Rev. Lett.* **65** 353–6
- [109] Esters M, Oses C, Hicks D, Mehl M J, Jahnátek M, Hossain M D, Maria J-P, Brenner D W, Toher C and Curtarolo S 2021 Settling the matter of the role of vibrations in the stability of high-entropy carbides *Nat. Commun.* **12** 5747
- [110] Calzolari A, Oses C, Toher C, Esters M, Campilongo X, Stepanoff S P, Wolfe D E and Curtarolo S 2022 Plasmonic high-entropy carbides *Nat. Commun.* **13** 5993
- [111] van de Walle A, Tiwary P, de Jong M, Olmsted D L, Asta M, Dick A, Shin D, Wang Y, Chen L-Q and Liu Z-K 2013 Efficient stochastic generation of special quasirandom structures *Calphad* **42** 13–18
- [112] Bergerhoff G, Hundt R, Sievers R and Brown I D 1983 The inorganic crystal structure data base *J. Chem. Inf. Comput. Sci.* **23** 66–69
- [113] Bokas G B, Chen W, Hilhorst A, Jacques P J, Gorsse S and Hautier G 2021 Unveiling the thermodynamic driving forces for high entropy alloys formation through big data *ab initio* analysis *Scr. Mater.* **202** 114000
- [114] Mehl M J, Hicks D, Toher C, Levy O, Hanson R M, Hart G L W and Curtarolo S 2017 The AFLOW library of crystallographic prototypes: part 1 *Comput. Mater. Sci.* **136** S1–S828
- [115] Hicks D, Mehl M J, Gossett E, Toher C, Levy O, Hanson R M, Hart G L W and Curtarolo S 2019 The AFLOW library of crystallographic prototypes: part 2 *Comput. Mater. Sci.* **161** S1–S1011
- [116] Hicks D, Mehl M J, Esters M, Oses C, Levy O, Hart G L W, Toher C and Curtarolo S 2021 The AFLOW library of crystallographic prototypes: part 3 *Comput. Mater. Sci.* **199** 110450
- [117] Kumaran S V, Garbiec D and Torralba J M 2023 A novel and sustainable method to develop non-equatomic CoCrFeNiMo_x high entropy alloys via spark plasma sintering using commercial commodity powders and evaluation of its mechanical behaviour *Mater. Sci. Eng. A* **878** 145207
- [118] Shi Y, Collins L, Feng R, Zhang C, Balke N, Liaw P K and Yang B 2018 Homogenization of Al_xCoCrFeNi high-entropy alloys with improved corrosion resistance *Corros. Sci.* **133** 120–31
- [119] Otto F *et al* 2016 Decomposition of the single-phase high-entropy alloy CrMnFeCoNi after prolonged anneals at intermediate temperatures *Acta Mater.* **112** 40–52
- [120] Lin S F, Pierce D T and Spicer W E 1971 Photoemission studies of platinum *Phys. Rev. B* **4** 326–9
- [121] Fadley C S and Shirley D A 1970 Electronic densities of states from x-ray photoelectron spectroscopy *J. Res. Natl. Bur. Stand. A* **74A** 543–58
- [122] Rempp H 1974 On the electronic density of states of the transition metals: VI. Isochromat spectroscopy of the 5d-metals iridium, platinum and gold *Z. Phys.* **267** 181–6
- [123] Furmanchuk A, Saal J E, Doak J W, Olson G B, Choudhary A and Agrawal A 2018 Prediction of Seebeck coefficient for compounds without restriction to fixed stoichiometry: a machine learning approach *J. Comput. Chem.* **39** 191–202
- [124] Takeuchi A and Inoue A 2001 Quantitative evaluation of critical cooling rate for metallic glasses *Mater. Sci. Eng. A* **304–306** 446–51