Accelerated chemical space search using a quantum-inspired cluster expansion approach

Chemical space search is a challenging problem, given the large size of the chemical space and complex structure-property relationship. In this work, we propose a solution by mapping the task of optimizing material property to that of finding the ground state of an Ising model. In our benchmarking, we observe significant time acceleration using a quantum-inspired device. Applying this to discovery of catalyst, we discover a new stable and efficient quaternary oxide catalyst for acidic oxygen evolution reaction.

**Highlights**
- Mapping chemical space search to Ising model enables accurate and efficient search
- Significant speed-up over genetic algorithm, bayesian optimization, and commercial solver
- Development of electronic structure proxy to predict in-laboratory catalyst performance
- Discovery of RuCrMnSbO₂, which is stable and efficient for acidic oxygen evolution

Choubisa et al., Matter 6, 1–21
February 1, 2023 © 2022 Elsevier Inc.
https://doi.org/10.1016/j.matt.2022.11.031
Accelerated chemical space search using a quantum-inspired cluster expansion approach

Hitarth Choubisa,1,10 Jehad Abed,1,10 Douglas Mendoza,2,3 Hidetoshi Matsumura,4 Masahiko Sugimura,4 Zhenpeng Yao,2,3,6 Ziyun Wang,1,5 Brandon R. Sutherland,1 Alán Aspuru-Guzik,6,7,8,9,* and Edward H. Sargent1,11,*

SUMMARY
To enable the accelerated discovery of materials with desirable properties, it is critical to develop accurate and efficient search algorithms. Quantum annealers and similar quantum-inspired optimizers have the potential to provide accelerated computation for certain combinatorial optimization challenges. However, they have not been exploited for materials discovery because of the absence of compatible optimization mapping methods. Here, by combining cluster expansion with a quantum-inspired superposition technique, we lever quantum annealers in chemical space exploration for the first time. This approach enables us to accelerate the search of materials with desirable properties 10–50 times faster than genetic algorithms and bayesian optimizations, with a significant improvement in ground state prediction accuracy. We apply this to the discovery of acidic oxygen evolution reaction catalysts and find a promising previously unexplored chemical family of Ru-Cr-Mn-Sb-O2. The best catalyst shows a mass activity eight times higher than state-of-the-art RuO2 and maintains performance for 180 h.

INTRODUCTION
Finding materials with desirable properties such as high d-band centers, high stability, high mobilities, optimal bandgaps, or low overpotentials to enable efficient energy harvesting,1 catalysts,2 light emission,3 sensing,4 and more is a challenging high-order optimization problem. Traditionally, density functional theory (DFT)-based screening5,6 has been used to explore chemical space. However, the high computational cost associated with DFT calculations7 and the vastness of chemical space8 make exhaustive searches infeasible. To predict material properties rapidly, studies in recent years have employed machine learning (ML)-based surrogate models.9–13 Among them, approaches based on learning from stoichiometries14,15 and generative models such as variational autoencoders (VAEs)16–18 and generative adversarial networks (GANs)19–22 allow accurate property prediction without DFT relaxation, which is a computationally expensive step. However, despite the presence of an accurate ML model, chemical exploration can still be limited because of the generation of infeasible structures by generative models, absence of efficient search strategies, and large chemical space (see Note S1 for an estimate). Therefore, ensuring global optimization across an exhaustive search within chemical space is difficult with such methods. This is further complicated by the observation that formulating the chemical space search as an optimization problem is itself challenging due to the absence of a simple analytic optimization expression and the non-convex nature of structure-property relationships due to the presence of multiple local minima.23,24

PROGRESS AND POTENTIAL
Searching the chemical space exhaustively is difficult because of its large size and non-convex structure-property relationship, preventing discovery of optimal alloys and doped materials. Here, we develop a method that maps the search for materials with optimal properties to that of finding the ground state of an Ising model. This enables the use of specialized combinatorial optimization devices such as the quantum annealers and quantum-inspired optimizers, resulting in an efficient and accurate search. We apply this to the discovery of acidic oxygen evolution reaction catalysts and discover a previously unreported stable and efficient catalyst demonstrated through in-laboratory experiments. We envision that this approach will be used to further explore diverse and more complex alloyed structures that were previously difficult to explore because of computational complexity and that this method will provide a way to test usefulness of various computing architectures to demonstrate computational advantage.
We aim to address this challenge by developing a new method to search chemical space by reframing the problem as one of finding the Ising Hamiltonian’s ground state (Figure 1). Our method maps the cluster expansion (CE) formulation to an Ising model. This mapping enables potential use of quantum annealers and quadratic unconstrained binary optimization (QUBO) solvers to perform rapid global searches in chemical space. We choose a quantum-inspired optimization engine (Digital Annealer [DA]; refer to the methods section “DA” for more details) as the choice to solve for the ground state of our Ising model. Our choice is inspired by the fact that DA has been shown to efficiently find ground states of Ising models. We term the approach quantum-inspired CE (QCE). When used on its own, QCE allows us to circumvent the DFT relaxation process and simultaneously enable us to accelerate the search for materials that optimize target properties of interest (Figure 1). At the same time, QCE can also be used in conjunction with superior property prediction models such as interatomic potentials that can accurately predict the properties to exhaustively search the material space and find materials with desirable properties.

In this paper, we first detail the QCE method and benchmark it by finding stable materials in the quaternary Cu-Ni-Pd-Ag chemical system. Following the benchmark, we leverage the framework to find stable and efficient acidic oxygen evolution reaction (OER) catalysts by developing an experimentally verified quantum mechanical catalyst efficiency proxy. Finally, we verify our predictions experimentally and our top catalyst (a never-before-reported multi-metal oxide) demonstrates best-in-class stability among all acidic rutile OER catalysts with overpotentials <250 mV.

RESULTS
Mapping CE to Ising Hamiltonian instance
C Es are power series expansions of the partition function that account for many-body interactions. They have been widely employed for the exploration of materials. In CE, the desired property P is expressed as a function of the vector σ, represents the type of atom at a specific lattice site. N is the total number of sites in the lattice. The property of interest P is expressed as a function of the vector σ:

\[ P(\sigma) = \sum J_s \Pi(s) \]  (Equation 1)

where \( J_s \) is the fitted correlation coefficient and \( \Pi(s) \) is the correlation function of multi-body interaction terms \( \sigma \). By expanding the correlation functions in Equation 1, we get:

\[ P(\sigma) = \sum_{i \in \text{cluster}_1} J_i(\Phi(\sigma),\Phi(\sigma)) + \sum_{i \in \text{cluster}_2} J_i(\Phi(\sigma) \cdot \Phi(\sigma)) + \sum_{i,j \in \text{cluster}_3} J_i(\Phi(\sigma) \cdot \Phi(\sigma) \cdot \Phi(\sigma)) \]  (Equation 2)

where \( \Phi(\sigma) \) is an orthogonal basis function dependent on the atom type \( \sigma \). Terms \( \text{cluster}_1 \), \( \text{cluster}_2 \), and \( \text{cluster}_3 \) are single-body, two-body, and three-body interactions of a unique cluster of sites. The \( J_s \) coefficients are obtained by fitting the CE on the DFT-generated data (initial structures and relaxed structure properties) (Figure S1B). The CE is an infinite series, but it can be approximated as a finite series.

The key step of our proposed mapping is expressing the basis functions \( \phi(\sigma) \) as a superposition of all basis functions for every site within the lattice:
such that $c(M_j)$ is the linear combination coefficient where the summation of the coefficients is equal to 1, $M_j$ is one chemical element from the periodic table, and $j$ represents the index of that element ($j = 1, \ldots, n$). It is due to the use of this superposition with constraints on coefficients that we name our method QCE.
In this study, we use the encoding \( c(M_j) = q_j \), where \( q_j \) is a binary variable only accepting 0 and 1 as the values. Such variables are quite often realized as spins of a quantum annealer or binary decision variables in quantum-inspired optimizers. Several other candidate functions can also be used for the encoding scheme \( c(\cdot) \) (refer to the methods section “Encoding schemes for QCE” for an in-depth analysis of alternative encoding schemes). Using the encoding \( c(M_j) = q_j \) transforms Equation 3 as

\[
\Phi_i(\sigma_i) = \sum_{j=1}^q q_j \varphi_i(\sigma_i = M_j)
\]  

(Equation 4)

Substituting for \( \Phi_i \) from Equation 4 in Equation 2 yields an expression with several multi-body interaction terms \( (q_0q_1, q_0q_1q_2, q_0q_1q_2q_3, \ldots) \). We use term reduction as reported by Babbush et al.\(^4\) to simplify the expression repeatedly until a QUBO expression is obtained:

\[
P = \sum_{i,j=1}^N A_{ij}q_iq_j
\]  

(Equation 5)

where \( A_{ij} \) are QUBO coefficients that are a function of \( J_i, I_{ij} \) and penalties used for term reduction. \( q_i \) is a binary decision variable accepting only 0 or 1 as the value. \( N \) is the total number of decision variables including the auxiliary variables used for term reduction.

The QUBO expression is then transferred to DA for efficient search and finding configurations that have the optimal property (minimum or maximum). Since any QUBO expression can be mapped to an Ising Hamiltonian, our proposed approach can be used by any Ising model solver such as a quantum annealer for optimizing materials within chemical space. This further motivated the term “QCE.” The complete process is summarized in Figure 1.

### Benchmarking our approach through exploration of quaternary chemical space (Cu-Ni-Ag-Pd)

To benchmark the performance of our proposed approach, we explored the quaternary chemical space spanned by Cu-Ni-Ag-Pd in a face-centered cubic (FCC) lattice due to their potential applications as fuel cell membranes\(^4\) and oxygen reduction electrocatalysts.\(^4\) Stability of materials for such applications is critical and is being extensively investigated.\(^4,43\) Thus, finding stable alloys within this chemical space can be an exciting prospect\(^41,44\) (Figure 2A). We use DFT for data generation and perform structure relaxation on randomly decorated FCC structures, followed by evaluation of the total energy of the systems. These total energies are then used to calculate mixing energies \( (\Delta H_{mix}) \), which are used to train a CE model (refer to the sections “Data generation” and “Mixing energy \( (\Delta H_{mix}) \)” for more details) and use it as a measure of alloy’s stability against its single metal constituents. The model achieves a 10-fold cross-validation mean absolute error (MAE) of 8 meV/atom (Figure 2B; see methods section “CE training” for more details). We used CE with two-body and three-body interactions with cutoff radius of 12 Å as we found it to be most accurate in terms of cross-validation mean absolute error.

Using the scheme presented in the previous section, we prepared the QUBO representation of our CE model and transferred it to the DA for finding structures with minimum mixing energy (refer to the methods section “DA” for more details). To explore the potential of our approach, we sought to determine stable alloys within the Cu-Ni-Pd chemical subspace. We achieved this by restricting the possible values
of the basis function $\Phi_i$ to only include relevant elements in the expansion of Equation 3 (refer to the methods section “Chemical subspace search” for more details). Upon performing the search using DA, we obtained equimolar ordered $\text{Cu}_{0.5}\text{Pd}_{0.5}$ alloy as the alloy with most negative mixing energy ($\Delta H_{\text{mix}} = -0.20$ eV/atom) in Cu-Ni-Pd space and, therefore, it is stable (Figure S2). This observation is independently supported by a previous report about equimolar FCC Cu-Pd alloy being the more stable.\(^{44}\)

Furthermore, we compared the temporal and efficiency performance of our approach to other widely used search algorithms in chemistry and material science: genetic algorithms (GAs) and bayesian optimization (BO). We implement the GAs through distributed evolutionary algorithms in Python (DEAP) framework\(^ {45}\) and BO using scikit-optimize\(^ {46}\) (for exact implementation, please refer to our code).
We find that our approach when used in combination with the DA significantly outperforms the alternatives in finding the minimum mixing energy structure (Figure 2C; logarithmic scale). We also observe that our approach results in better quality of global minima. Even though the energy difference between QCE-based predictions and best GA-based predictions is 0.03 eV/atom, the ground state structures found by them are significantly different (Cu₀.₅Pd₀.₅ and Cu₀.₆₉Pd₀.₃₁ respectively) indicating the importance of accurate search algorithms. In addition, when we compare the results of DA against BO, we observe much worse performance than DA as well as GA. The final compositions found by BO were much closer to the starting points, indicating that BO used in our comparisons is more effective in searching for local optimum than global optimum. All the tests were performed using a 20 core Intel Xeon 6148 CPU operating at 2.4 GHz. Detailed parameters for all the comparison algorithms are presented in the methodssection “Performance benchmark algorithms.”

**Finding stable and efficient OER materials**

To demonstrate the generalizability of our approach, we showcase the discovery of acidic OER catalysts. The development of OER catalysts has been the subject of significant research over the past few years. For the alkaline medium, several stable and efficient OER catalysts have been proposed; however, only a few stable, efficient, and cost-effective OER catalysts have been reported in literature for operation in an acidic medium.

High efficiency and high stability are two crucial requirements for a good OER catalyst. To develop a quantitative model for efficiency of the catalyst candidates, we built upon the d-band model by Hammer and Nørskov and recent studies about its interplay with p-band centers. We used materials and alloys constructed by substituting two different prototype structures: ZrO₂ (P2₁/c; monoclinic) and RuO₂ (P4₂/mmm; tetragonal) in our study. We generate slabs to model the catalytic activities for randomly generated structures (refer to methods section “Data generation”). After performing DFT calculations, we used the adsorption energies to determine the theoretical overpotentials (refer to Figure S4 for model of the slabs). We performed logistic regression on the differential improvement of overpotential; i.e., if overpotential of one material is smaller than the other (refer to methods section “Logistic regression analysis”; Figure 3A). We observe that the probability of material i having an overpotential smaller than materials j is given by the following expression:

$$\text{Prob}(\eta_i < \eta_j) = \frac{1}{1 + e^{\theta}}$$

(Equation 6)

where \(\theta = 0.078 \cdot (E_p - E_d) - 0.165 \cdot (E_d - E_d)\)

where \(\eta\) represents overpotential and \(E_p\) and \(E_d\) represent p-band center and d-band center, respectively.

To further establish the validity of this theoretical analysis, we further conduct several in-laboratory experiments. Using just the contributions due to band centers, we achieve a Pearson’s correlation coefficient of 0.67 (\(p < 0.001\)) between experimental overpotentials and theoretical proxy. We also observe that, when combined with enthalpy of mixing (\(\Delta H_{\text{mix}}\) per atom), we observe a significant improvement in predictive power using simple Ridge regression. Inclusion of mixing energy as a predictor significantly improves 10-fold cross-validation correlation from 0.67 to 0.81 (Figures 3B and 3C; Note S3). This role of mixing energy as a predictor can
Figure 3. Development of an efficiency proxy for OER catalysts and training the CEs

(A) The relationship between electronic structure of the material and its efficiency expressed through overpotential. We trained a logistic regression-based classifier to assess whether one material has a larger overpotential than the other (0 if smaller and 1 if larger) and use it to search for more efficient materials.

(B) We further validate this hypothesis on overpotentials obtained through in-laboratory experiments as part of this study; when represented as band centers, we are able to cluster the differences with high accuracy.

(C) How the theoretically calculated d- and p-band centers correlate with experimental overpotentials. We use the exact coefficients obtained through logistic regression comparison classifier to perform this analysis.

(D–F) The performance of CE models used for exploration within this study for mixing energies.

(D) Predictions of the bulk structures for [Ru-Cr-Mn-Sb-Ti-V-W-Co]$_2$O$_2$.

(E) Performance of CE model in predicting d-band centers of the bulk structures for [Ru-Cr-Mn-Sb-Ti-V-W-Co]$_2$O$_2$.

(F) Performance of CE model in predicting p-band centers of the bulk structures for [Ru-Cr-Mn-Sb-Ti-V-W-Co]$_2$O$_2$. 

Please cite this article in press as: Choubisa et al., Accelerated chemical space search using a quantum-inspired cluster expansion approach, Matter (2022), https://doi.org/10.1016/j.matt.2022.11.031
be explained as stable catalysts are likely to catalyze the reaction consistently. The overall expression to predict overpotentials $\eta$ is expressed as,

$$\eta = K(0.046E_p - 0.097E_d - \Delta H_{mix} + C) \quad \text{(Equation 7)}$$

where $E_p$ is p-band center, $E_d$ is the d-band center, $\Delta H_{mix}$ is mixing energy in eV/atom, $C$ is a constant, and $K$ is just a multiplier. The values for the three properties are not on the same scale and that explains the large variation in the absolute value of coefficients. For exact values of $K$ and $C$, please refer to Note S3. We also report all the collected experimental quantities for the candidates with measurable activity and acidic stability in Table S3 as a tabular database for future ML explorations and for reference. In addition, we also used mixing energy $\Delta H_{mix}$ to quantify stability of an alloy relative to its precursors (i.e., ease of synthesizability of the alloy).

Following the strategy for QCE, we trained CE models (Figure S5; methods section “CE training”) and generated the QUBO representations for each of the three properties, $\Delta H_{mix}$ ($H_E$), p-band center ($H_p$), and d-band center ($H_d$), using QCE. These properties were obtained by first performing DFT geometry optimization on the initial decorated lattices followed by calculation of the converged electronic and thermodynamic properties. Performing cross-validation analysis to find the optimal cutoffs, we find that CE models with 4 and 8 Å cutoffs for two-body and three-body interaction terms lead to the best predictive power. Similarly, for predicting d-band centers, the highest predictive power is demonstrated by a CE with 12 and 6 Å cutoffs for two-body and three-body interactions. A CE model that incorporated just two-body interactions with a cutoff of 4 Å achieved the best predictive power. The accuracy and performance of each of these models on validation data are shown in Figures 3D–3F. We defined QUBO for efficiency as,

$$H_{eff} = 0.046E_p - 0.097E_d - \Delta H_{mix} \quad \text{(Equation 8)}$$

Thus, the problem of finding an efficient and stable OER catalyst transforms into finding a state that minimizes both $H_{eff}$ and $H_{stable}$ where $H_{stable}$ is the QUBO representation for mixing energies. We solve this problem by using a heuristic approach of optimizing the linear combination of the two QUBOs:

$$H_{mixed} = \lambda_1 H_{stable} + \lambda_2 H_{eff} \quad \text{(Equation 9)}$$

where $\lambda_1, \lambda_2 \in [0, \infty)$ (refer to Figure S6 for the effect of parameters $\lambda_1, \lambda_2$). Transferring $H_{mixed}$ represented as QUBO to DA and performing the optimization result in candidates that have promising stability as well as efficiency.

We searched for different chemical subspaces within the two prototype structures by varying the relative weight parameters $\lambda_1, \lambda_2$ and chemical subspace of interest (Figure S7) while adding constraints to ratios of different elements (refer to Figure 4 for a summary of candidates experimentally tested and Table S1 for a detailed list), whereas non-zero pairs lead to a balanced trade-off between stability and efficiency (Figure S6). We also compare this approach with a sequential search (search for stability followed by efficiency). We find that our heuristic approach outperforms a sequential search strategy (Note S7). In our searches, we focused on exploring ternary and quaternary chemical spaces. This choice was motivated by existing exploration of binary transition metal oxides in the literature as well as effectiveness of the sol-gel method of synthesis for multi-metal oxide synthesis.

In the context of the presented QCE method for chemical space search, it is also worth noting the previous studies by Pedersen et al. on usage of BO for catalyst search. Both the observations made by the authors of that study and our
benchmarking experiments indicate that a BO-based search strategy is better suited to a local optimum search than a global optimum. So, if the question is to find a globally optimal material, QCE will outperform BO-based approaches. On the other hand, both BO and QCE with constraints can be used for local optimizations of materials within the chemical space. With QCE, we can constraint the elemental fractions as inequalities and perform local chemical space search.

Figure 4. Experimental screening of electrocatalysts
(A) The mixing energy of all electrocatalyst candidates considered for experiments in this study. The candidates were classified based on the synthesis outcome: synthesis failed (the electrocatalyst was not produced), synthesized, and pure rutile phase.
(B) Overpotential screening of synthesized electrocatalysts.
(C) The intrinsic activity of electrocatalysts normalized by the electrochemical surface area.
(D) The mass activity of electrocatalysts normalized by the total mass of Ru.
(E and F) (E) XRD and (F) polarization curves (Tafel slope analysis in the inset) of the best electrocatalyst candidate R14-M10-41 (Ru0.58Cr0.25Mn0.083Sb0.083O2) and a baseline RuO2 catalyst.
(G) Chronopotentiometry test of the electrocatalysts sprayed on carbon paper electrodes. In the inset is the accelerated test conducted on the electrocatalysts drop-casted on carbon paper. The electrocatalyst loading for all samples is 1 mg/cm2. The electrolyte is 0.5 M H2SO4.
Experimental verification

All of the 49 electrocatalyst candidates in Table S1 were synthesized using sol-gel synthesis followed by annealing in air for 2 h.\(^61\) We observed that some compositions (eight compounds) were not successfully synthesized using the sol-gel method due to poor promotion of networks in the gel (Figure 4A). The crystal structure of the synthesized samples was screened using X-ray diffraction (XRD). XRD patterns were collected and classified into two groups: pure rutile phase (21 compounds) and multiple-phases groups (20 compounds). One XRD pattern is shown in Figure 4E for Ru\(_{0.58}\)Cr\(_{0.25}\)Mn\(_{0.08}\)Sb\(_{0.08}\)O\(_2\) (R14-M10-46) showing characteristic peaks of a rutile structure similar to a baseline RuO\(_2\) catalyst. The XRD peaks are broad, indicating that our synthesis approach produced nanocrystalline electrocatalysts. Further analysis of the structure for all electrocatalysts is provided in Table S2. High-resolution transmission electron microscopy (HR-TEM) and scanning transmission electron microscopy (STEM) are shown for three samples in Figure S8. The nanoparticles have an average particle size of ~10 nm with a spherical shape suitable for electrocatalytic applications. The energy X-ray dispersive spectroscopy (EDS) elemental mapping shows a homogeneous distribution of elements in the three samples (Figure S9). The EDS elemental mapping of Ru\(_{0.58}\)Cr\(_{0.25}\)Mn\(_{0.08}\)Sb\(_{0.08}\)O\(_2\) (R14-M10-46) matches within 5% error with the nominal composition of the compound (Ru, 59 ± 5.9 atom %; Cr, 20 ± 1.0 atom %; Mn, 9 ± 0.6 atom %; Sb, 13 ± 3.4 atom % of the total metal amount). The electron diffraction rings of the three samples confirm the formation of a strained rutile structure (Figure S10).

The electrocatalysts were then drop-casted on carbon paper to screen the electrochemical performance. The overpotential of the electrocatalysts was measured in a three-electrode cell using 0.5 M H\(_2\)SO\(_4\) (Figure 4B; Table S3). The experimental overpotentials matched well with the trend of predicted activity (Figure 3C; Note S3). We observe a coefficient of determination of 0.71 between our theoretically derived quantum mechanical proxy and in-laboratory experimental overpotential of catalyst candidates (\(p < 0.001\)). We used the structures as obtained by optimization of Equation 9 to model every chemical composition. Furthermore, to compare the intrinsic activity of the electrocatalysts, we decoupled the morphology effect from the activity by normalizing the current density by the active electrochemical surface area overpotential instead of the geometric area (Figures 4C and S11). We then calculated the mass activity of the electrocatalysts normalized by the Ru amount in the electrocatalyst (Figure 4D). Taking into consideration both measures, we identified Ru\(_{0.58}\)Cr\(_{0.25}\)Mn\(_{0.08}\)Sb\(_{0.08}\)O\(_2\) as the most promising candidate with the highest mass activity 381 A/g\(_{Ru}\) (about eight times higher than RuO\(_2\)) and a much lower overpotential increase rate of 2 mV h\(^{-1}\), 10 times slower than RuO\(_2\) (inset in Figure 4G; Table S3).\(^62\) To better assess the stability of the electrocatalyst, we prepared it by spraying the electrocatalyst on carbon paper instead of drop-casting, the technique we used for screening purposes, to ensure a higher penetration of the particles through the hydrophilic carbon paper yielding in higher surface coverage and better mechanical attachment to the carbon fibers. We maintained the loading of the electrocatalyst between both preparation techniques at 1 mg/cm\(^2\). Our candidate maintained an overpotential of less than 300 mV for 180 h at a low overpotential increase rate of 2 mV/h, 10 times lower than RuO\(_2\) and any previous reports of rutile structure-based catalysts.\(^43\) Finally, we also search existing databases (OQMD, AFLLOW, Materials Project, OCP, ICSD) and literature for similar alloys and we were unable to find even a similar oxide compound with Ru, Cr, Mn and Sb present together. Our QCE method was therefore able to effectively search chemical space, identify a new promising family of multi-metal oxides for exploration, and discover a highly efficient and stable OER catalyst.
We also compared the results with a random sampling of chemical composition strategy. We observed that five out of eight of our predictions demonstrated an overpotential of less than 300 mV, and three of them demonstrated an exceptional stability of less than 10 mV/h overpotential degradation (Table S3). If we had to randomly sample the chemical space to find at least one of these three compositions, we would need to perform at least ~766 experiments in the quaternary chemical space of just a single chemical subspace (see Note S8 for stoichiometric space size), indicating at least 95× improvement over a random chemical space search strategy purely based on experiments.

**DISCUSSION**

**Origins of the catalyst stability and efficiency**

To better understand the cause of higher activity and stability of our top catalyst candidate, Ru$_{0.58}$Cr$_{0.25}$Mn$_{0.09}$Sb$_{0.08}$O$_2$ (henceforth referred to as catalyst RuCrSbMnO$_2$), we performed detailed DFT calculations and analysis. We used the crystal structure for each candidate, as obtained by minimizing the joint Hamiltonian in Equation 9, for performing the following analysis (Figures S12 and S13 for crystal structure of RuCrMnSbO$_2$ and its XRD comparison to experimental structure). Bader charge analysis showed that the average partial charge on Ru increased from 1.73|e| in pure RuO$_2$ to 1.82|e| in RuCrMnSbO$_2$. Ru with a higher partial charge indicates improved ability of the catalyst to oxidize water to oxygen (i.e., OER). Additionally, to explore the origin of the stability of the catalyst candidates, we further analyzed the density of states (DOS). As shown in Figures 5A and 5B, incorporation of Cr, Mn, and Sb alter the DOS of RuO$_2$, causing occupation at Fermi level to decrease from 38 states/spin · cell to 18 states/spin · cell. This decrease in DOS at fermi energy indicates stronger bonding between metal atoms, leading to stabilization of our solid solution (see Figure S14 for a comparison study that considers only Cr and Mn). Table 1 summarizes our observations against previously reported Ru$_{0.4}$Cr$_{0.6}$O$_2$ and RuO$_2$. We also observe that entropic contributions further stabilize the predicted composition RuCrMnSbO$_2$ compared with previously reported RuO$_2$ and Ru$_{0.4}$Cr$_{0.6}$O$_2$. These observations align with our observed experimental measurements of 180-h stability for RuCrMnSbO$_2$ compared with unstable RuCrO$_2$ and 40-h stability of pure RuO$_2$ (catalyst ID R8-M16-16 and RuO$_2$ for detailed experimental results on stability and activity in Tables S1–S3).

At the same time, we calculated free-energy profiles ($\Delta G$) of OER to compare the activities of RuCrMnSbO$_2$ with RuO$_2$ (refer to methods section "Data generation" for details and Notes S5 and S6). As can be seen from Figure 5 (C and D), the formation of OOH was found to be the rate-determining step. For RuCrMnSbO$_2$, the Bader charge was found to be 1.92 eV, which is smaller than the Bader charge corresponding to RuO$_2$ (2.02 eV) and consistent with the experimental observation of smaller overpotentials. This observation can be explained by the finding that the d-band surface DOS is higher for RuCrMnSbO$_2$ (7.52 states/spin · cell) compared with RuO$_2$ (5.30 states/spin · cell), leading to the lower overpotential and lower reaction energy barrier to the formation of adsorbed OOH radical. Presence of larger number of states enables easier electron transfer and, therefore, facilitates formation of intermediates (*O, *OH, *OOH).

**Conclusions**

In this study, we reported a new approach for materials discovery that maps the chemical space search problem to one of finding the ground state of the Ising Hamiltonian. This enables us to use a quantum-inspired computing framework to find materials with optimal properties in an accelerated and efficient fashion, orders of magnitude faster than the widely used alternatives, such as GAs and BO. Our efforts led us to develop an improved efficiency proxy and discover a stable and efficient
OER catalyst \( \text{Ru}_{0.58}\text{Cr}_{0.25}\text{Mn}_{0.09}\text{Sb}_{0.08}\text{O}_2 \). Post hoc DFT analysis further explains the electronic origins of the stability and efficiency of the catalysts. However, we should also take note of one of the limitations of the current study: restriction of one lattice across the chemical space, which is not what is always desired in materials discovery pipelines. Nevertheless, this can be handled by using extended CE mapping as proposed by Koretaka Yuge.\(^{65}\) We plan to incorporate this as part of our future work.

**EXPERIMENTAL PROCEDURES**

**Resource availability**

**Lead contact**

The lead contact for this paper is Edward H. Sargent (ted.sargent@utoronto.ca).

---

**Figure 5.** Post hoc DFT analysis of the best candidates

(A) Bulk DOS decomposed into elemental contributions and normalized to one for all the constituent elements for RuO\(_2\).

(B) Elemental decomposed DOS for bulk RuCrMnSbO\(_2\) using the same normalization scale. We observe that the DOS for RuO\(_2\) at fermi level is 38 states/(eV\(\cdot\)spin) larger than 18 states/(eV\(\cdot\)spin) in RuCrMnSbO\(_2\), which is an indicator of better chemical bonding.\(^{63}\)

(C) Free-energy diagram for OER for RuO\(_2\) catalyst.

(D) Free energy diagram for OER with RuCrMnSbO\(_2\) catalyst. In both cases, \(^*\)OOH formation is the rate-determining step (RDS) with \(\eta_{\text{RuO}_2} > \eta_{\text{RuCrMnSbO}_2}\) where \(\eta\) represents overpotential.

OER catalyst \( \text{Ru}_{0.58}\text{Cr}_{0.25}\text{Mn}_{0.09}\text{Sb}_{0.08}\text{O}_2 \). Post hoc DFT analysis further explains the electronic origins of the stability and efficiency of the catalysts.
### Materials availability

This study did not generate new unique reagents.

### Data and code availability

All the code and data supporting this study are available at a public Github repository: [https://github.com/hitarth64/quanum-inspired-cluster-expansion](https://github.com/hitarth64/quanum-inspired-cluster-expansion). The repository contains all the code required to run the searches and data used for bulk periodic systems within the study.

All the slab calculation data are available at: [https://github.com/hitarth64/MixedMetalOxides](https://github.com/hitarth64/MixedMetalOxides).

### Methods

#### DA

Fujitsu’s DA is designed to efficiently solve combinatorial optimization problems \(^2^7\) formulated as fully connected Ising problems expressed in QUBO form. In the most general case, a QUBO can be represented as

\[
F(q_0, q_1, \ldots, q_n) = \sum_{ij} a_{ij} q_i q_j,
\]

where \(q_i \ (i = 1 \ldots n)\) are binary variables.

In this work, we used a third-generation DA, which can handle up to 100,000 binary decision variables. Given a QUBO expression, the DA finds the assignment of binary variables (referred by \(q_0, q_1, \ldots\)) such that the expression value is minimized. All experiments were conducted on the DA environment prepared for research use. More details on exact algorithms and configuration of DA can be found at: [https://www.fujitsu.com/jp/documents/digitalannealer/researcharticles/DA_WP_EN_20210922.pdf](https://www.fujitsu.com/jp/documents/digitalannealer/researcharticles/DA_WP_EN_20210922.pdf).

#### Data generation

We use RuO\(_2\) (Materials Project\(^{6^6}\): mp-825) and ZrO\(_2\) (Materials Project: mp-2858) as the prototype structures and generated 72 and 96 atom supercells, respectively, for training the CE models. Then, 120 randomly substituted alloy structures were generated for each of the chemical spaces spanned by Ru-Ti-W-Sb-Cr-Mn-V-Co and Ru-Zr-Hf-Y-V-Co-Fe-Ce. We further randomly selected 34 structures and generated surfaces with 110 orientations for the calculation of adsorption energies with vacuum spacing of 10 Å on both sides. We chose 110 orientations since we found it to be most stable with respect to surface energy (refer to **Note S6**). The slab models were composed of 144 atoms (48 metal atoms and 96 oxygen atoms) repeated in four layers besides the adsorbates (O, OH, and OOH). DFT calculations were performed in Vienna ab initio simulation package (VASP) for the bulk and surface structures using Perdew-Burke-Ernzerhof\(^{6^7}\) (PBE) exchange-correlation functional augmented with Hubbard coulomb interaction potential (U) corrections for d-electrons taken from Materials Project.\(^{6^6}\) Valence electrons were described with a 520-eV plane-wave basis set and 0.05 eV Gaussian smearing of the electronic density. All the bulk geometries were optimized with the energy convergence criterion of \(10^{-4}\) eV and force convergence of 0.03 eV/ Å. Core electrons were described using the projector augmented wave (PAW) method. All the calculations were spin polarized and reciprocal space was simulated using a \(3 \times 2 \times 1\) k-points mesh.
centered at gamma. Refer to Note S2 for a detailed evaluation of the computational costs associated with a single calculation.

For Cu-Pd-Ni-Ag chemical space benchmark, we used Cu-FCC 64 atoms supercell as the prototype structure, and 120 randomly decorated structures were generated. DFT calculations were performed in VASP for the bulk using Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional. Valence electrons were described with a 520-eV plane-wave basis set and 0.05 eV Gaussian smearing of the electronic density. All the geometries were optimized with the energy convergence criterion of $10^{-4}$ eV and force convergence of $10^{-3}$ eV/A. Core electrons were described using the PAW method.

To perform post hoc DFT analysis of the RuO$_2$ and RuCrMnSbO$_2$, we generated surfaces with orientations [001, 100, 101, 110] and choose 110 as the orientation to perform adsorption energy calculations due to lower energy (Note S5). The post hoc DFT analysis was done on a k-points grid $3 \times 2 \times 1$ with 520 eV as the energy cutoff for the PAW pseudopotentials and strongly constrained and appropriately normed (SCAN) exchange-correlation functional was used with the Hubbard-U correction. The rest of the parameters stayed the same as in the preceding paragraph.

**Mixing energy ($\Delta H_{\text{mix}}$)**

We define mixing energy or enthalpy of mixing $\Delta H_{\text{mix}}$ as excess energy with respect to the single metal precursors. For metal alloys it assumes the form

$$\Delta H_{\text{mix}}(A_xB_yC_zD_w) = E(A_xB_yC_zD_w) - \frac{x \cdot E(A_x) + y \cdot E(B_y) + z \cdot E(C_z) + w \cdot E(D_w)}{u}$$

where $u = x + y + z + w$ and $E(\cdot)$ represents total energy of the concerned phase.

Similarly, for mixed metal oxides, it assumes the form

$$\Delta H_{\text{mix}}(A_xB_yC_zD_wO_v) = E(A_xB_yC_zD_wO_v) - \frac{x \cdot E(A_xO_v) + y \cdot E(B_yO_v) + z \cdot E(C_zO_v) + w \cdot E(D_wO_v)}{u}$$

where $u = x + y + z + w$.

**CE training**

We used integrated cluster expansion toolkit (ICET) to train CE models. The scripts and codes are available at https://github.com/hitarth64/quantum-inspired-cluster-expansion. We tested different cluster generation schemes by tuning the kind of interactions and cutoff for each of the interactions (see Figure S5 for a comparison plot). Parameters that lead to the lowest cross-validation error were chosen for analysis.

**Encoding schemes for QCE**

Several candidate functions can be used for the encoding scheme $C(\cdot)$. Two natural choices are as follows:

- **Encoding 1:**
  $$\Phi_i(\sigma_i) = \sum_{j=1}^{n} a_j \cdot \varphi_i(\sigma_i = M_j)$$
where \( q_j \) are binary decision variables and the one-hot constraint of Equation 3 is then enforced as a penalty:

\[
P_i(\sigma) = \left( \sum_{j=1}^{n} q_j - 1 \right)^2
\]

- Encoding 2:

\[
\Phi_i(\sigma) = \sum_{j=1}^{n} \Phi(q_1, q_2, \ldots, q_k) \cdot \varphi(\sigma_i = M_j)
\]

where \( q_j \) are binary decision variables, \( k = \text{ceil}(\log_2 n) \) where \( n \) is total number of element types. Functions \( \Phi_i(\cdot) \)'s are defined such that only one \( \Phi_i(\cdot) \) is non-zero and others are zero for any \( \vec{q} \in \{0,1\}^k \) and each \( \Phi_i(\cdot) \) is one at least for one \( \vec{q} \in \{0,1\}^k \). Here are the steps to construct such \( \Phi_i \). First, for each possible bit string in the space \( \{0,1\}^k \), construct a polynomial that takes one only for the bit string and zero for other bit strings. Next, determine a surjection from the bit strings to element indices \( \{1, \ldots, n\} \). Finally, define \( \Phi_i \) as the sum of the polynomials correspond to bit strings that mapped to the element \( j \) in the previously defined surjection. Such an encoding scheme trivially enforces the one-hot constraint of Equation 3 and therefore, we do not need to add it as a penalty. For a three elements system example, the bit strings and the corresponding polynomials are \{0,0\} and \( \Phi(1 - q_1) \cdot \Phi(1 - q_2) \), \{0,1\} and \( \Phi(q_1) \cdot \Phi(1 - q_2) \), \{1,0\} and \( \Phi(1 - q_1) \cdot \Phi(q_2) \), and \{1,1\} and \( \Phi(q_1) \cdot \Phi(q_2) \).

Performance benchmark algorithms

We performed comparison against four heuristic search algorithms that are widely used in chemistry and material science. Their implementation details are as follows:

- GA: GA is implemented using the DEAP framework. \(^{45}\) We use eaSimple algorithm with 300 generations and a population size of 300 to run and perform the searches. We use crossover and mutation probabilities of 0.3 and 0.7.
- GAMuPlusLambda: this search method uses \((\mu + \lambda)\) evolutionary algorithm to perform the chemical space search using eaMuPlusLambda implementation as provided in DEAP. We use 300 generations and a population size of 300 with crossover and mutation probabilities the same as above. Corresponding \( \mu \) and \( \lambda \) parameters that yielded the optimal solution were found to be 300 and 400.
- BO: we conducted BO-based searches to search and optimize the alloy combinations using scikit-optimize. \(^{46}\) We specifically used Gaussian processes combined with a hybrid acquisition function that randomly chooses either lower confidence bound (LCB), expected improvement (EI), or probability of improvement (PI) at every iteration.
- Gurobi: we used commercial Gurobi 9.5.2 solver \(^{69}\) and solved the chemical space optimization as a quadratic integer program. Gurobi is available free of charge for academic purposes.

The exact values are tabulated in Table 2.

Chemical subspace search

For performing QCE, we know from Equation 3 that we rewrite the CE formulation as
\[ F_i(\sigma) = \sum_{j=1}^{n} C(M_j) \cdot \varphi_i(\sigma_i = M_j) \]

such that \( M_j, j = 1 \ldots n \) represent the set of elements being considered for the site \( i \) and

\[ C(M_j) \in \{0, 1\} \quad \forall \ j \]
\[ \sum_{j=1}^{n} C(M_j) = 1 \]

In order to restrict the sampling to a set of elements, we limit the summation to set of elements of interest to us. If one wants to avoid element \( M_s \), one can rewrite the summation as,

\[ F_i(\sigma) = \sum_{j=1, j \neq s}^{n} C(M_j) \cdot \varphi_i(\sigma_i = M_j) \]

Such that \( C(M_j) \in \{0, 1\} \quad \forall \ j \)
\[ \sum_{j=1}^{n} C(M_j) = 1 \]

OER mechanism

OER in acidic and neutral medium is a four-step reaction with the following reaction mechanism:\(^5,43,50\):

\[ \star + H_2O \leftrightarrow OH^+ + (H^+ + e^-) \]
\[ OH^+ \leftrightarrow O^\bullet + (H^+ + e^-) \]
\[ O^\bullet + H_2O \leftrightarrow OO^\bullet H^+ + (H^+ + e^-) \]
\[ OO^\bullet H^+ \leftrightarrow \star + O_2 + (H^+ + e^-) \]

where \( \star \) denotes the species adsorbed at the catalyst surface.

The theoretical overpotential (\( \eta_{OER} \)) is thus defined as

\[ \eta_{OER} = \max \left( \Delta G_7, \Delta G_8, \Delta G_9, \Delta G_{10} \right) - 1.23 \ V \]
\[ = \max \frac{\Delta G_{OH^+} - \Delta G_{OOH^+} - \Delta G_{OH^+} - \Delta G_{OOH^+}}{e} - 1.23 \ V \]

Calculation of \( \eta_{OER} \), thus, requires values of adsorption energies for reaction intermediates \( O, OH, \) and \( OOH \). These DFT energy calculations are computationally expensive. Usually, scaling relationships are used to reduce the number of calculations needed.\(^{47}\) However, scaling relationships break down in the case of alloys. Therefore, here, we perform explicit DFT calculations for all three adsorbate structures, \( \star O, \star OH, \) and \( \star OOH \), along with bare surface to calculate the adsorption.
energies. These adsorption energies along with zero point energies and entropic corrections are then used to calculate $\Delta G_{O/H}$, $\Delta G_{O}$, and $\Delta G_{OOH}$ as

$$
\Delta G_{O/H} = E_{DFT}(O^\ast) - E_{DFT}(O) - E_{ZPE}(O) + T \Delta S(O)
$$
$$
\Delta G_{O} = E_{DFT}(OH^\ast) - E_{DFT}(OH) - E_{ZPE}(OH) + T \Delta S(OH)
$$
$$
\Delta G_{OOH} = E_{DFT}(OOH^\ast) - E_{DFT}(OOH) - E_{ZPE}(OOH) + T \Delta S(OOH)
$$

Furthermore, to overcome the limitation of traditional GGA-based calculations for O$_2$ molecule, its free energy is calculated as

$$
G(O_2) = 4.92 + 2G(H_2O) - 2G(H_2) \text{ (in eV)}
$$

### Calculation of band centers

D-band model relates the surface reactivity with the shifts in the d-band center of the catalyst where the d-band center is a single state with energy $\varepsilon_d$ that approximates the interaction of participating bands of d-states. In particular, we calculated d-band centers and p-band centers using the first moments as

$$
\varepsilon_{M-d} = \frac{\int_d E_{M-d} (E - E_F) \, dE}{\int_d 0_{M-d} (E - E_F) \, dE}
$$
(Equation 10)

$$
\varepsilon_{M-p} = \frac{\int_p E_{M-p} (E - E_F) \, dE}{\int_p 0_{M-p} (E - E_F) \, dE}
$$
(Equation 11)

where M is the corresponding element and $D_{M-y}$ represents partial DOS of orbital y of element M in the system.

### Logistic regression analysis

We used logistic regression, as implemented in scikit-learn, to investigate the relationship between band centers and overpotentials of the materials. We first calculated theoretical overpotentials (section “OER mechanism”) of 34 randomly chosen candidate materials from the two chemical families under consideration (section “Data generation”). We trained the classifier to distinguish the differential improvement between two catalysts; i.e., trained to predict whether $\eta_i < \eta_j$ (overpotential of material i is less than that of material j). In addition, refer to Figure S3 to see the variation in intermediate adsorption energies as a function of band centers.

### Materials

Metal precursors including ruthenium chloride hydrate (RuCl$_3$·xH$_2$O), chromium(iii) chloride anhydrous, 99.99% trace metals basis (CrCl$_3$), manganese chloride (MnCl$_2$), tungsten (VI) chloride (WCl$_6$), titanium diisopropoxide bis(acetylacetonate), 75 wt % in isopropanol, vanadium(iii) chloride (VCl$_3$), and propylene oxide (≥ 99.5% GC) were all purchased from Sigma-Aldrich. Nafion (5 wt % in a mixture of lower aliphatic alcohols and water) and AvCarb MGL190 were used for electrode preparation. All chemicals were used without any further purification.

### Experimental synthesis

The mixed metal oxides were synthesized using a modified sol-gel procedure$^{61}$ by completely dissolving a total amount of 2.7 mmol of the metal precursor mixture in 3.5 mL of anhydrous ethanol. The solution was vortexed for 5 min and sonicated in a water bath for 1 h until it was clear. Then, it was chilled in a refrigerator for 2 h to prevent any undesired hydrolysis and condensation, which may affect the gelation process. Afterward, a magnetic stirrer was used to mix the solution vigorously while 2 mL of propylene oxide was added dropwise to the mixture. The solution was aged for 1 day to promote gel formation and then washed with acetone; the process was repeated for 5 days before drying the gel in a vacuum oven for 1 day. The sol-gel
process produces amorphous oxides; therefore, to produce a rutile crystal structure, the dried powder was annealed at 400°C for 2 h.

For fast electrochemical purposes, the working electrode was prepared on 0.5 × 0.5-cm untreated carbon paper (AvCarb MGL190) by drop-casting. First, a catalyst ink was prepared by mixing 10 mg of the catalyst in 1 mL of a mixture of water and ethanol (4:1, v/v). Then, the ink was sonicated in an ice bath for 1 h. Finally, a 25-μL drop of the ink was deposited on the carbon paper and allowed to dry in the air. For final electrochemical testing, the working electrode was prepared by spraying the electrocatalyst using N2 on 0.5 × 0.5-cm untreated carbon paper. The ink was prepared by mixing 10 mg of the catalyst in 1 mL of isopropanol. Then, the ink solution was sonicated in an ice bath for 1 h before spraying it on the carbon paper on the hot plate at 90°C.

**Electrochemical testing**

Electrochemical testing was conducted in a three-electrode setup using a Hg/Hg2SO4 as the reference electrode and graphite rod as the counter electrode. The OER activity was evaluated by running linear sweep voltammetry (LSV) with a rate of 5 mV/s. The stability of the catalyst was evaluated by conducting chronopotentiometry at 10 mA/cm².

All potentials were iR-corrected by measuring the solution resistance from electrochemical impedance spectroscopy (EIS) with a bias of 1.53 V vs. reversible hydrogen electrode (RHE) in the frequency range from 100 kHz to 10 MHz and an amplitude of 5 mV. All the potentials in this study were reported with respect to an RHE using the following relationship:

\[ E_{\text{RHE}} = E_{\text{Hg/Hg2SO4}} + 0.640 + 0.0591 \times \text{pH} \]

**Structure analysis**

The crystal structure of catalysts was determined using X-ray diffraction (XRD). A Miniflex 600 (Rigaku, Japan) equipped with D/teX Ultra silicon trip detector and Cu Kα radiation (λ = 1.5418 Å) was used. Powders were prepared by mixing with acetone and then dropping a small drop of the mixture to fill a 4-mm-diameter by 100-μm-deep groove in a single crystal silicon holder (zero-background). The angle was varied between 20° and 80° with a step size of 0.005° and a scan rate of 1°/min.

To evaluate the nanocrystalline size and macrostrain of the electrocatalysts, we used Scherrer’s equation and microstrain equation:

\[ D = \frac{K \lambda}{\text{FWHM} \cos \theta} \]

Scherrer’s equation  
\[ \epsilon = \frac{\text{FWHM}}{4 \tan \theta} \]

microstrain  

D is the mean crystallite size, λ is 0.154 nm for Cu X-ray source, K is shape factor and has a typical value between 0.9 and 1. The full-width half-maximum (FWHM) of the peaks were measured by fitting to Gaussian distribution and then calculated from the standard deviation (σ) using the following relation:

\[ \text{FWHM} = 2\sqrt{2 \ln 2} \approx 2.355 \sigma \]

**Electron microscopy**

The structural characterization and elemental mapping of the catalysts were done using high-resolution scanning electron microscopy (HRSEM) and TEM. The experiments were conducted in a Hitachi HF3300 equipped with a cold field emission
electron gun using an accelerating voltage of 300 kV. An EDS detector was used in STEM mode to analyze and quantify the composition of the nanoparticles. Also, a secondary electron (SE) detector was used to collect HRSEM images of the nanoparticles. Powder samples were prepared in ethanol and sonicated for 10 min before drop-casting a 1- to 2-μL drop on a 400-mesh copper grid and drying overnight.

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.matt.2022.11.031.

ACKNOWLEDGMENTS
This work was financially supported by Fujitsu Ltd. and Fujitsu Consulting (Canada) Inc. We thank Fujitsu for their generous access to the Digital Annealer where all the annealing operations and optimizations were performed. All the DFT calculations were performed on Compute Canada’s Niagara supercomputing cluster. The experimental work was supported financially by the Natural Sciences and Engineering Research Council of Canada (NSERC), Vanier Canada Graduate Scholarship. Theoretical work was supported through Hatch Scholarship for Sustainable Energy Research. Electron microscopy was conducted at the Ontario Center for the Characterization of Advanced Materials (OCCAM).

AUTHOR CONTRIBUTIONS

DECLARATION OF INTERESTS
The authors declare no competing interests.

REFERENCES


