



Discovery of complex oxides via automated experiments and data science

Lusann Yang^a, Joel A. Haber^b, Zan Armstrong^a, Samuel J. Yang^a, Kevin Kan^b, Lan Zhou^b, Matthias H. Richter^b, Christopher Roat^a, Nicholas Wagner^a, Marc Coram^a, Marc Berndt^a, Patrick Riley^a, and John M. Gregoire^{b,1}

^aGoogle Research, Google Applied Science, Mountain View, CA, 94043; and ^bDivision of Engineering and Applied Science and Joint Center for Artificial Photosynthesis, California Institute of Technology, Pasadena, CA 91125

Edited by Alexis T. Bell, University of California, Berkeley, CA, and approved August 2, 2021 (received for review April 16, 2021)

The quest to identify materials with tailored properties is increasingly expanding into high-order composition spaces, with a corresponding combinatorial explosion in the number of candidate materials. A key challenge is to discover regions in composition space where materials have novel properties. Traditional predictive models for material properties are not accurate enough to guide the search. Herein, we use high-throughput measurements of optical properties to identify novel regions in three-cation metal oxide composition spaces by identifying compositions whose optical trends cannot be explained by simple phase mixtures. We screen 376,752 distinct compositions from 108 three-cation oxide systems based on the cation elements Mg, Fe, Co, Ni, Cu, Y, In, Sn, Ce, and Ta. Data models for candidate phase diagrams and three-cation compositions with emergent optical properties guide the discovery of materials with complex phase-dependent properties, as demonstrated by the discovery of a Co-Ta-Sn substitutional alloy oxide with tunable transparency, catalytic activity, and stability in strong acid electrolytes. These results required close coupling of data validation to experiment design to generate a reliable end-to-end high-throughput workflow for accelerating scientific discovery.

data science | materials discovery | complex oxides | optical absorption | oxygen evolution electrocatalysis

Increased incorporation of data science in materials research is anticipated to accelerate discovery of materials with improved properties and combinations thereof for technological applications requiring multifunctional materials (1, 2). Machine learning is one popular approach for building predictive models, but limited materials training data often compromises the prediction accuracy, especially in composition spaces for which no training data are available (3–5). Training data are particularly limited in high-order composition spaces (e.g., at least three cation oxides), which offer opportunities for tuning multiple properties through formation of a phase, i.e., a crystal structure or substitutional alloy, that contains all three cations. The vast number of potential high-order compositions exceeds current methods of discovery or prediction (6–9), and prediction of substitutional alloy phases and their properties remains a substantial challenge (10, 11).

We develop two data science methods to discover materials in high-order composition spaces. The phase diagram model uses thermodynamic equilibrium assumptions to propose candidate phase diagrams using only optical absorption data. The emergent property model uses the same data to identify compositions whose optical properties cannot be explained by combinations of lower-order compositions of the same elements. The present work additionally describes the design and implementation of the high-throughput workflow that provides data to these models as well as an example use case for guiding discovery. Our primary finding is that appropriately constructed data science models can make inferences about the phase behavior of complex materials using data that are not traditionally used for phase characterization. These inferences add scientific value to existing datasets and guide materials discovery efforts.

We demonstrate this approach for three-cation oxide systems via high-throughput experiments coupled to automated quality control and modeling of spectral microscopy data. The select three-cation oxide compositions whose properties appear unique compared to lower-order oxide compositions are then candidates for more expensive and time-consuming structural and functional characterization. This approach is distinct from computational inverse design wherein a model predicts a material to have a specific property, a promising strategy that is hampered by the dual challenges of computational prediction of experimental properties and the computational generation of synthesizable materials (12, 13). Our approach shifts the strategy from identifying materials with a specific property to rapidly screening materials that may be exceptional for any property. By releasing the database of experiments and analyses alongside this work, we aim to accelerate the community's selection of composition spaces and compositions therein for discovery of materials exhibiting a broad range of properties (14).

Discovering complex phases with desirable properties, whether by experiment or computation, is highly challenging due to the combinatorics of composition spaces. Searching the Materials Project (15) for entries containing oxygen, having an associated Inorganic Crystal Structure Database entry (16), having unique composition and space group, and excluding inert gas and nonmetallic elements

Significance

Automation is accelerating the discovery of useful materials, yet testing even a small fraction of the billions of possible materials for a desired property is beyond the reach of workflows involving resource-intensive property measurements. Due to relationships among composition, structure, and properties, identifying a complex material with one interesting property makes it the proverbial needle in a haystack that merits testing for additional properties. We accelerate materials synthesis and optical characterization by employing physics-aware data science to identify materials for further investigation. With this approach, one does not need high-throughput methods for measuring every material property of interest since a single ultra-high-throughput workflow can guide material selection for other properties, which is a new paradigm for accelerated materials discovery.

Author contributions: L.Y., J.A.H., M.B., P.R., and J.M.G. designed research; L.Y., J.A.H., K.K., L.Z., and M.H.R. performed research; L.Y., J.A.H., Z.A., S.J.Y., C.R., M.C., M.B., P.R., and J.M.G. contributed new reagents/analytic tools; L.Y., J.A.H., N.W., and J.M.G. analyzed data; and L.Y., J.A.H., and J.M.G. wrote the paper.

Competing interest statement: As listed in the affiliations, several authors are current or former employees of Google, a technology company that sells machine learning services as part of its business.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: gregoire@caltech.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2106042118/-DCSupplemental>.

Published September 10, 2021.

(He, Ne, Ar, Kr, Xe, Rn, C, N, F, P, S, Cl, Se, Br, I, and H) yields 755 1-cation oxide entries from 73 cation elements. Applying the same search to two-cation oxides increases the number of identified materials to 4,345, although the corresponding search for three-cation oxides yields only 3,163 materials. While some two-cation oxide phases undoubtedly remain to be discovered, there has been extensive computational exploration of two-cation oxides, making such materials the focus of recent high-throughput computational (17–19) and machine learning–driven materials discovery (20–23). Higher-order composition spaces enable further tuning of materials properties, but the expense of comprehensive search of combinatorial spaces is clear when considering 3-cation oxides. Using the 73 cation elements from the 1-cation oxide entries, there are 62,196 (73 choose 3) possible 3-cation oxide composition spaces, yet only 2,205 are represented in the Materials Project, leaving over 96% of the composition spaces with no existing data.

The computational exploration of three-cation phases to date has focused on crystal structures where each cation has a unique crystallographic site. The site substitution of multiple elements on a single crystallographic site is a distinguishing feature of metal substitutional alloys, and a metal oxide structure exhibiting such substitutions on cation sites is referred to herein as a substitutional alloy oxide (or “alloy” for brevity). A three-cation oxide can crystallize in a structure observed in the one or two-cation subspaces, and the composition-tuned decoration of the cation sublattice comprises an opportunity for tuning properties in the three-cation composition space. Since the site substitution is disordered, large unit cells combined with ensemble averaging of different random site decorations are required to explicitly model substitutional alloys. While approximations to computational modeling of alloys have been developed (24–27), alloys in high-order composition space comprise a dramatically underexplored class of materials for discovery efforts. We know from the examples of high-temperature superconductors and catalysis that extremely valuable properties are obtainable via substitutional alloying in high-order composition spaces (8, 28, 29).

We report a high-throughput workflow for discovering candidate compositions for functional properties by coupling high-throughput synthesis and optical characterization with automated data interpretation. Parallel optical screening was recently demonstrated as a proxy for phase behavior in the context of combinatorial thermal processing of individual compositions (30). We extend this approach to high-order composition spaces

using inkjet printing (31) to deposit composition-gradient lines of material that are subsequently imaged by a purpose-built hyper-spectral microscope that measures optical absorption from the infrared to ultraviolet (UV). We present a dataset consisting of nine channels of optical absorption data for a series of metal oxide composition samples. Each composition sample is defined by the stoichiometry of cation elements, with oxygen content driven toward equilibrium by calcination at fixed oxygen pressure. The dataset contains 376,752 distinct compositions from 108 three-cation oxide systems based on the cation elements Mg, Fe, Co, Ni, Cu, Y, In, Sn, Ce, and Ta, for which only the Ce-Cu-Fe oxide system contains an entry in the Materials Project. We present a data science workflow incorporating cross-validation and other quality control measures to establish confidence in the data, enabling subsequent data modeling to predict aspects of the underlying phase behavior. In the present work, we discuss models that 1) predict candidate phase diagrams along with the absorption spectrum of each phase (the “phase diagram model”) and 2) predict the likelihood that the three-cation composition space contains a three-cation phase whose properties are distinct from one or two-cation phases (the “emergent property model”). These complementary prediction models are emblematic of the usage of data from high-throughput experiments to make inferences that accelerate resource-intensive experiments.

This implementation of data science–driven analysis of experimental data are complementary to quantum mechanical (32) and machine learning (22, 33) prediction of new phases. Detection of interesting systems and compositions via the modeling of optical data can seed an investigation for new phases and/or for determining whether the three-cation compositions exhibit exceptional properties. Our approach builds upon a foundation of combinatorial materials science in which synthesis of composition libraries is coupled to measurement of properties of interest (34–45). While providing a direct route to discovery of a desirable property in a specific composition system, this approach limits exploration of many composition spaces due to both the high relative expense of property measurements and the need to measure every composition library for every property of interest. Modeling phase behavior from optical properties to guide further experiments is illustrated herein for the Co-Ta-Sn oxide composition space, for which X-ray diffraction (XRD) experiments verified the discovery of the (Sn,Co,Ta)₂O₂ rutile substitutional alloy oxide. Furthermore, screening of the alloy compositions for electrocatalysis of the oxygen evolution

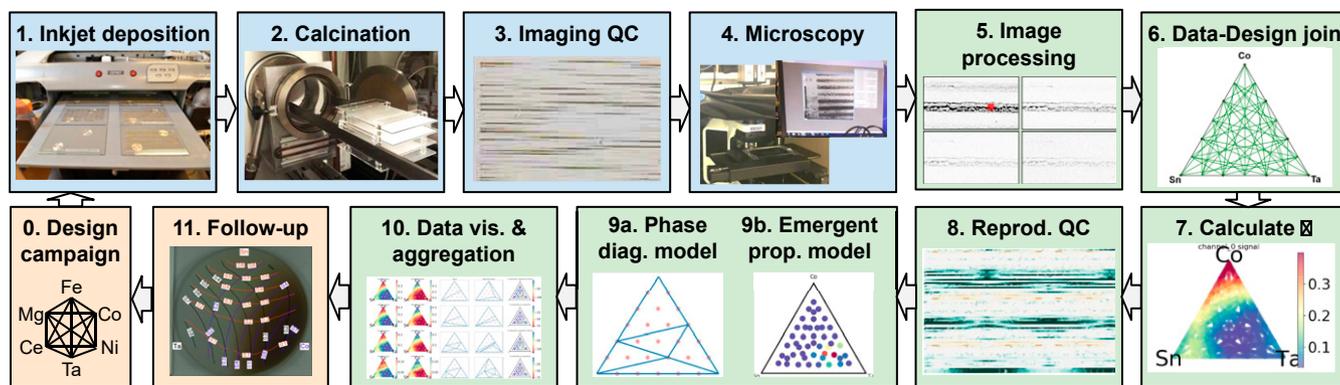


Fig. 1. Workflow for synthesis, characterization, and analysis of metal oxide libraries. One iteration through the workflow involves a batch of 20 composition libraries corresponding to the 20 three-cation composition spaces for a set of six elements. Each experimental process (blue, steps 1 to 4) is parallelized for 1 to 20 plates at a time. Each analysis step (green, steps 5 to 10) is computationally automated with manual quality control. Given a selection of cation elements (step 0), three-cation composition libraries are deposited (step 1) and calcined (step 2). Imaging of each library plate for quality control (step 3) is followed by high-resolution spectral microscopy (step 4). Image processing (step 5) and identification of the locations of each printed composition (step 6) enable modeling of composition-dependent spectral absorption (step 7). Provided sufficient reproducibility within the composition library (step 8), candidate phase diagrams and compositions exhibiting emergent optical properties (step 9) are presented to users (step 10) to design follow-up measurements (step 11).

reaction revealed an optimal combination of activity and stability, which was enabled by the optical discovery despite the lack of any explicit relationship between the optical and catalytic properties. Although demonstrated for three-cation oxides, the methodology is designed to be implementable in even higher-order composition spaces.

Results

To establish a high-throughput workflow for exploring three-cation oxide composition spaces, we carried out a suite of viability and feasibility experiments. A primary result of the experiment-computation iterative design was the workflow (Fig. 1) that integrates highly parallelized experimental processes with automated computational processes, whose outcomes inform follow-up experiments as well as the choice of elements and conditions to reseed the workflow.

The materials synthesis portion of the workflow commences with deposition of metal oxide precursor solutions onto aluminosilicate glass library plates using an inkjet printer (46). For process optimization, we designed the composition library using discrete lines of material aligned with the fast-raster direction of the print head, with each line containing a continuous composition gradient. Our focus on the development of procedures based on linear composition gradients was also motivated by the transferability of the workflow to higher-order composition spaces. As described in greater detail (*SI Appendix, section A*), a linear composition gradient spanning a composition space with N elements can intersect $(N-2)$ -dimensional phase field boundaries such that all phase boundaries are detectable via analysis of a sufficiently dense set of composition lines. For the three-cation composition space, we used a network of 45 composition lines spanning the ternary composition triangle, the composition graph for three-cation oxides with oxygen stoichiometry determined by materials processing (*SI Appendix, Fig. S1C*). Combined with the two-cation composition lines at the perimeter of the ternary composition triangle, this corresponds to 48 total composition line segments (see *SI Appendix, Figs. S2–S4*).

A final primary design choice of the synthesis workflow is the materials scope of a single print session. We balanced practical considerations of printer operation with the desire to synthesize broad swaths of three-cation composition space, leading to the choice that each print session includes a set of six cation elements (Fig. 1, step 0). All 20 possible combinations of three metals from the six inks are deposited, one combination per library plate (*SI Appendix, section A*). Following inkjet printing (Fig. 1, step 1), the set of plates from a print session undergo parallel drying and calcination processes, culminating with tube furnace calcination at 750 °C for 10 h (Fig. 1, step 2).

Each plate is imaged with a flatbed scanner to aid visual inspection (Fig. 1, step 3), which at a minimum confirms that material was deposited across each plate and that deposited material is constrained to the intended lines as opposed to diffuse deposition indicative of print head failure. The library plates are then characterized using hyperspectral microscopy (Fig. 1, step 4), the final high-throughput experiment. Data analysis commences with image processing (Fig. 1, step 5) to obtain spectral transmittance images that are subsequently aligned to the intended deposition pattern (Fig. 1, step 6).

To mitigate the influence of the thickness fluctuations, which are inherent to the inkjet-printed materials, the optical absorption, i.e., a spectral absorption coefficient (α) with arbitrary units, is calculated using a nonnegative matrix factorization algorithm (Fig. 1, step 7). The α signal from each composition sample is then compared to the median value from 10 composition duplicates deposited on the same plate, enabling visual quality control for reproducibility (Fig. 1, step 8). At this point in the workflow, the dataset for each composition line segment within each three-cation library plate consists of the 10-fold replicate α signals, the

aggregated signal α_{comp} , and the uncertainty σ_{α} for each of the nine optical channels (photon energies).

While this rich dataset is amenable to a host of analyses, the present work focuses on two complementary data interpretation models (Fig. 1, step 9). Both models rely on an underlying assumption that the thin-film compositions equilibrate with the O_2 atmosphere during the 10 h anneal at 750 °C and that kinetic limitations prohibit phase changes during cooling and storage at ambient temperature (*SI Appendix, section C*). It is worth noting that annealing at higher temperatures may exceed liquidus temperatures of many compositions, which would complicate the thermodynamic equilibrium assumption, because such compositions would crystallize at different temperatures during cooling. Annealing at lower temperatures may limit diffusion and thus limit equilibration with the O_2 in the anneal atmosphere. Collectively, these limitations mean that it would be nontrivial to adapt the methods of the present work to include the oxygen composition and temperature axes in phase diagram modeling. For the primary goal of guiding materials discovery efforts with data models, processing very thin films at a sufficiently high temperature to promote equilibration enables data modeling under this plausible thermodynamic assumption.

In the phase diagram model, we approximate a phase diagram that is consistent with the α data. We consider a discretization of the ternary composition space with composition intervals of 1/6 (composition intervals of circa 16.7 at.%), yielding 15 and 10 composition points with two and three cations, respectively. For a phase diagram with K phases in addition to the three end-members, there are $25 \text{ choose-} K$ placements of the phase points, which we exhaustively search. Various combinations of Alkemade lines for connecting the $K + 3$ phase points to form compatibility triangles are considered via a formation energy sampling scheme, and we assume that within each phase field the α signals vary linearly with composition (the lever rule). This approximation enables each candidate phase diagram to be fit to the measured α signals with a regression model containing a fit parameter for each photon energy at each of the $K + 3$ phase points. The phase diagram that best approximates all the α values, taking into account the observed noise, is determined for each value of K . The analysis thus provides a candidate phase diagram that best describes the composition map of optical absorption for each K .

Importantly, this approach to approximating the phase diagram does not explicitly model substitutional alloying, which would drastically increase the set of candidate phase diagrams, as described further in *SI Appendix, section C*. To ascertain the implications of this aspect of the phase diagram model, consider a two-cation phase A_{1-x}B_x , where a third cation can substitute for B up to a saturated alloy composition $\text{A}_{1-x}(\text{B}_{1-y}\text{C}_y)_x$. If the alloy obeys a linear rule-of-mixtures relationship between y and the measured properties, then the data will be well modeled by this phase diagram model with a fitted phase point at the saturated alloy composition. Multiple alloying degrees of freedom may require additional fitted phase points on the perimeter of the alloy phase field, and any nonlinear relationship between properties and alloy composition could result in fitted phase points within the alloy phase field to provide the best linear lever-rule approximation to the nonlinear composition-property data. Regardless, structure identification for both stoichiometric and alloy phases requires further characterization under the guidance of the fitted phase diagrams.

We use K_{1st3cat} as an integer metric to summarize the fitting results, which is the lowest value of K at which the phase diagram model results in a three-cation fitted phase point. A low value for this metric indicates that the composition trends within the three-cation composition space cannot be well described with only two-cation phases. However, higher values of this metric do not preclude the existence of a three-cation phase with spectral α that is distinct from all one and two-cation compositions because

high variability in α along the two-cation composition lines may drive the phase diagram model to fit the data using only two-cation phase points. We thus seek a metric to summarize whether optical properties indicative of nontrivial phase behavior exist within the three-cation space.

The emergent property model complements the phase diagram model by focusing on detection of emergent optical absorption within the three-cation composition space, optical properties that are inexplicable from the properties of one and two-cation materials. For this analysis, we discretize the composition space using composition intervals of 1/10 to define 30 composition regions on the perimeter of the ternary composition graph and 46 composition regions in the three-cation space. We then consider each of these 46 composition regions independently and apply Gibbs' phase rule per the thermodynamic equilibrium assumption noted previously, which requires that no more than three phases will coexist at a given three-cation composition. By considering each of the 4,060 (30 choose 3) combinations of signals from the subspaces, we identify the linear combination of signals that maximizes the likelihood given the measured three-cation signal and its variability from the three-cation composition region. The metric P is the probability of this maximum-likelihood-scenario signal, where lower P corresponds to stronger evidence of an emergent property in the three-cation space. The value of P characterizes each of the 46 composition regions, and taking the minimum value provides the metric for evaluating the existence of an emergent property upon combination of the three cations under consideration.

The results of these analyses are visualized alongside α_{comp} to create a graphical summary of each library plate (Fig. 1, step 10) to guide follow-up measurements (Fig. 1, step 11). Representative data renderings are shown (Fig. 2) for three different composition libraries, which were chosen to all be Fe-containing systems to demonstrate the variability in optical properties and composition trends through variation of the other two elements. A summary of the 108 three-cation composition libraries based on the phase diagram and emergent property models is shown in Fig. 2E. The three systems shown (Fig. 2 A–C) represent different regions of this summary plot.

The Fe-Co-Ta system is shown (Fig. 2A) as representative of a typical oxide library plate, being in the midrange of both $K_{1st3cat}$

and minimum P . The optical trends are largely smooth and monotonic in composition space. Comparison of the fitted phase diagrams at $K = 2$ and $K = 3$ (Fig. 2D) and comparison with the composition plot of P facilitates assessment of whether the data support the existence of a three-cation phase. In this case, there is no strong evidence for a three-cation phase because the P signal indicates that the composition regions around the candidate three-cation phase are well described by two-cation compositions, with the lowest probability composition region indicating that a small concentration of Fe alters the optical properties of Ta-Co oxides, which is likely due to a small degree of substitutional alloying. The conclusion is that this system is well described by two-cation phases being present on the Fe-Ta and Ta-Co composition lines, which is validated via XRD measurements (SI Appendix, Fig. S5). The identified two-cation phases are Ta_2CoO_6 and FeTaO_4 , which were previously known, with the three-cation compositions showing a mixture of these two phases, consistent with the candidate phase diagram with no three-cation phase points. XRD measurements on the thin metal oxide samples are often inconclusive due to the weak scattering signal. Indeed, our process optimization for rapid characterization and reproducibility of optical properties results in synthesis of samples that have an average thickness of ~ 10 nm (47), which, combined with typically poor crystallinity of the metal oxide samples, results in undetectable XRD signals for most of the compositions that do not contain the heaviest elements.

A system known to contain a three-cation phase was investigated during the process-development phase of the project, prior to the generation of 108 three-cation systems of the primary dataset. The optical and XRD characterization of the library of the Cu-Fe-In oxide system is shown in SI Appendix, Fig. S6. The $K = 3$ candidate phase diagram suggests the existence of a two-cation and a pair of three-cation phases. XRD validation characterized these three phases as CuFe_2O_4 , the target CuFeInO_4 phase, and possibly an alloy of this three-cation phase. The XRD results also indicate Fe substitution into CuO that is not detected with the phase-fitting analysis. Indeed, the analysis of α cannot detect all phases but rather indicates where phase behavior results in nontrivial alteration of optical properties.

Returning to our discussion of representative systems (Fig. 2), the Fe-Ni-In oxide system (Fig. 2B) is representative of a system with high $K_{1st3cat}$ and relatively high minimum P , corresponding

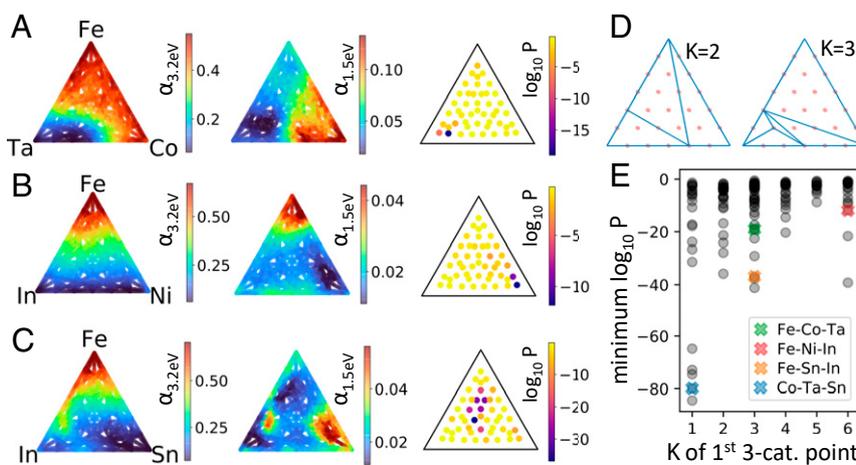


Fig. 2. Illustrative examples and summary of optical mapping of three-cation oxides. The composition maps of the absorption coefficient (α) at 3.2 eV and 1.5 eV as well as the results of the emergent property model ($\log_{10}P$) are shown for (A) Fe-Co-Ta, (B) Fe-Ni-In, and (C) Fe-Sn-In composition spaces. (D) Candidate phase diagrams with $K = 2$ and 3 fit points are shown for the Fe-Co-Ta system to illustrate the results of the phase diagram model. (E) The summary of 108 three-cation composition systems (gray points), including some duplicate systems from different print sessions. The horizontal axis is the lowest number of phase fit points (K) for which the fitted phase diagram includes a three-cation phase, and the vertical axis is the minimum log likelihood value ($\log_{10} P$) obtained from the 46 composition regions in the respective three-cation composition space. The four systems described in A to C as well as the Fe-Co-Ta system are denoted by colored markers.

to a system with no strong evidence of nontrivial three-cation phase behavior. Similar systems can be identified by considering three-cation systems in the summary figure (Fig. 2E) with $K_{1st3cat}$ of 5 or 6 and minimum $\log_{10} P$ above -30 . There are 28 such systems, corresponding to 21% of the library plates. Systems similar to the Fe-Co-Ta oxide system (Fig. 2A), where optically interesting two-cation compositions are apparent without strong evidence for nontrivial three-cation phases, can be identified as $K_{1st3cat}$ of 3 or 4 and minimum $\log_{10} P$ above -30 . There are 52 such systems, corresponding to 40% of the library plates. The other 51 systems, corresponding to 39% of the library plates, are the strongest candidates for identification of nontrivial three-cation phase behavior with respect to optical properties. This class of three-cation systems is less amenable to identification of a representative example, although one such system is the Fe-Sn-In oxide system (Fig. 2C) for which the three-cation composition trends, especially for $\alpha_{1.5eV}$, strongly indicate the presence of three-cation phases. Our inspection of the data from many of these systems along with select XRD measurements suggests that the three-cation phase behavior is often due to substitutional alloying within a two-cation phase, i.e., that all three cations coexist in a structure whose XRD pattern is indistinguishable from that of a known two-cation phase other than slight peak shifting that arises from alteration of the lattice constants.

Since three-cation substitutional alloying can introduce emergent properties in the three-cation space, these systems are of prime interest for further study, and in the present work, we describe one such follow-up study based on the Co-Ta-Sn oxide system (Fig. 3). This system has a relatively low minimum P and $K_{1st3cat} = 1$. Importantly, a $K_{1st3cat}$ value of 1 does not imply that the underlying phase diagram contains no two-cation phases, which would be unprecedented and counterintuitive to metal oxide chemistry. Rather, this result indicates that there are multiple phases in the system and that the most optically important composition trends for generating a candidate phase diagram occur within the three-cation space. The α data for this system appears unremarkable at first glance, but the emergent property analysis demonstrates that a region of Ta-rich compositions are more transparent across the full visible spectrum than any of the two-cation compositions. Given that the α values for

the Ta endmember, likely a Ta_2O_5 -type material, are already quite low, the lower α in three-cation space could be due to a change in refractive index that lowers the reflectivity. Regardless, the strong evidence of nontrivial phase behavior is due to a statistically significant change in optical properties that merits further investigation.

XRD analysis of select compositions within the library (SI Appendix, Fig. S7) demonstrates the presence of a phase in the three-cation space whose XRD pattern matches patterns for both Ta_2CoO_6 and TaO_2 rutile structures because the weak XRD signal is insufficient to distinguish these rutile structures. The Sn endmember appears to be rutile SnO_2 , prompting the design of a follow-up experiment to validate the presence of rutile alloys in the composition region bounded by these three known rutile phases. To synthesize a thicker material and validate the observation of nontrivial phase behavior with respect to optical properties, a continuous composition spread thin film (Fig. 4A) was sputter deposited and annealed using the same protocol as the printed library plate. The use of a different deposition method with the same annealing process also helps evaluate the thermodynamic assumptions noted previously. Subsequent XRD analysis reveals the presence of a rutile-type structure throughout the sputter-deposited region of the three-cation composition space, with a phase-pure rutile phase field straddled by phase fields where the rutile phase coexists with Co_3O_4 at Co-rich compositions and with Ta_2O_5 at Ta-rich compositions (Fig. 4B). The composition region of the rutile phase field and the continuous independent variation in lattice parameters indicate two compositional degrees of freedom of substitution on the cation sublattice. To our knowledge, a rutile substitutional alloy oxide containing these three cation elements has neither been experimentally observed nor computationally predicted, producing a successful demonstration of the discovery of a three-cation phase via analysis of optical properties. To validate the high-throughput optical measurements, the sputter-deposited composition spread was characterized by a combination of optical transmission-reflection spectroscopy measurements and X-ray fluorescence (XRF) measurements, providing the composition-dependent optical absorption (SI Appendix, Fig. S8), which is shown in Fig. 4C for the three photon energies of

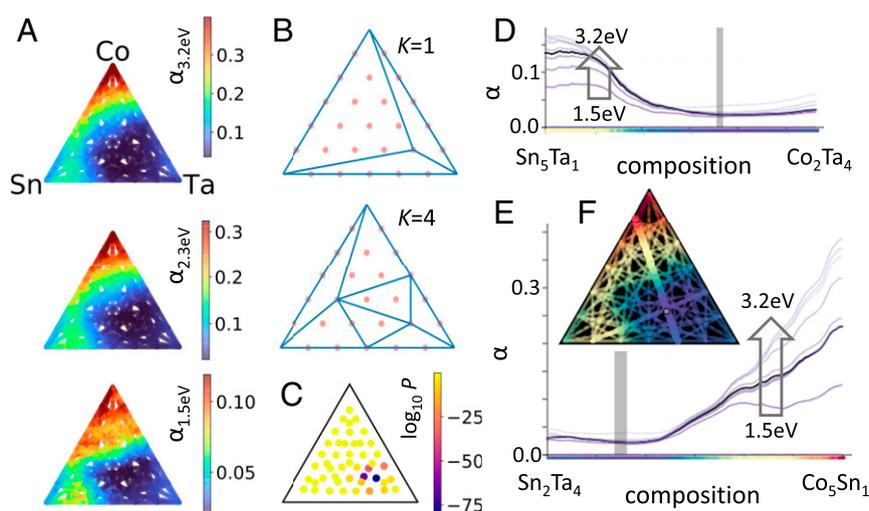


Fig. 3. Optical phase analysis of Sn-Co-Ta oxides. (A) Composition plots of the optical absorption signal at 3.2, 2.3, and 1.5 eV. (B) The $K = 1$ and $K = 4$ fitted phase diagrams indicating that the most optically important phase point is within the ternary composition space. (C) The log likelihood from the emergent property analysis, demonstrating that Ta-rich three-cation compositions exhibit optical properties that are markedly different from any combination of one and two-cation compositions (nonyellow points). (D and E) The nine-channel absorption signals, with the 2.0 eV signal shown with larger line width, are plotted along two composition lines, (D) Sn_5Ta_1 to Co_2Ta_4 and (E) Sn_2Ta_4 to Co_5Sn_1 . (F) The corresponding 2.0 eV signal is shown for all composition lines with the intersection at $Sn_{0.3}Co_{0.2}Ta_{0.5}$ denoted by a gray point, corresponding to the gray vertical region highlighted in D and E.

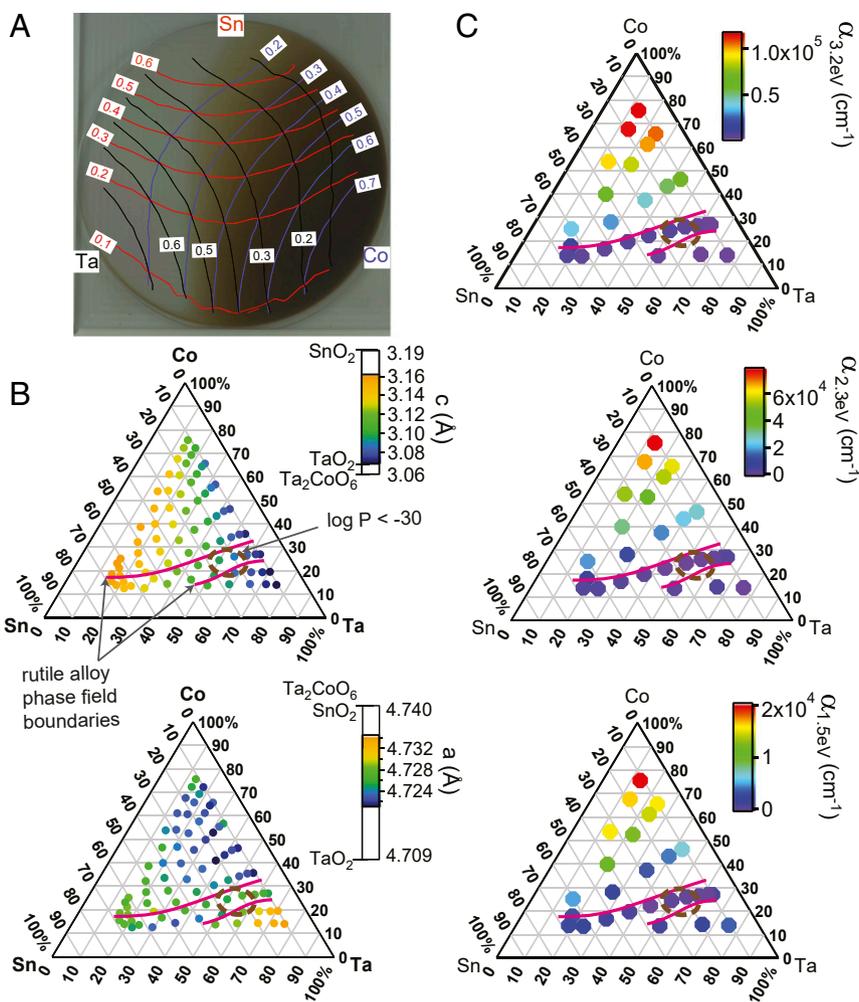


Fig. 4. Follow-up investigation of the Sn-Co-Ta oxide composition space using sputter-deposited thin films. (A) The photograph of the sputter-deposited film shows contour lines of cation composition determined by XRF. (B) The rutile unit lattice constants calculated from XRD measurements at 72 compositions are shown, where the color scales include annotations for the corresponding values for the previously known rutile phases in the system. Magenta lines indicate phase boundaries, where only the rutile phase was observed between the lines. The brown dashed line indicates the composition region from Fig. 3C with $\log P < -30$, which coincides with the rutile alloy phase field and validates the optical-based discovery of a three-cation phase. (C) For select compositions, optical absorption coefficients determined by transmission-reflection spectroscopy (at the same three photon energies as Fig. 3A) validate the low absorption in the composition region of interest. Note that the flat thin-film morphology of the sputter-deposited samples combined with film thickness measurements enables determination of the absorption coefficients in units of nm^{-1} .

Fig. 3. These data demonstrate excellent qualitative agreement with those of the high-throughput screening, most notably that the absorption in the three-cation rutile alloy region is quite low despite the presence of Co. The three-cation rutile oxides can exhibit similar or even lower absorption than more Ta-rich compositions, motivating future investigation for applications such as transparent conducting oxides (48).

In the *SI Appendix*, we provide additional follow-up experiments, including characterization of select compositions by spectroscopic ellipsometry (*SI Appendix*, Fig. S9). Following our assertion that emergent optical properties in three-cation space are a harbinger for other emergent properties, we continued follow-up measurements of this system by considering functional properties that have no known underlying relationship to optical absorption. Given the common use of Co in both metal and oxide catalysts and the use of Sn and Ta oxides as corrosion-resistant materials, we sought to ascertain whether the three-cation compositions offered a desirable combination of catalytic activity and stability against corrosion. Characterization of oxygen evolution electrocatalysis proceeded in pH 3 and then pH 0 (*SI Appendix*,

Fig. S10) electrolyte, highly corrosive conditions where Co oxides are known to suffer from rapid corrosion. These results indicate that Co is stabilized by its combination with Ta and Sn, providing a route to electrochemical stabilization of catalytic sites, as has been observed in other rutile substitutional alloy oxides (49). The additional compositional tuning of optical transparency is important for applications such as solar fuel generation, and collectively, the results demonstrate the three-cation tuning of multifunctional properties in a composition system identified by the computational analysis of optical data.

Discussion

The discovery of a composition-tuned family of electrocatalysts originated from modeling of optical microscopy data, highlighting the interrelationships of seemingly disparate materials properties through their underlying composition and structure, a central tenet of the materials exploration strategy described in the present work. Composition-structure-property relationships are ubiquitous in materials science, so identification of nontrivial composition-structure relationships can accelerate the identification of

nontrivial composition-property relationships for any property. Since high-throughput mapping of composition-structure relationships remains a challenge, especially in high-order composition spaces, we translate this problem into identification of nontrivial composition-optical property relationships to guide exploration of exceptional materials for other properties, which are often more difficult to measure than optical properties. Our strategy recognizes that most high-order composition spaces contain uninteresting mixtures of lower-order composition phases and that the functional properties of such high-order compositions may be unremarkable compared to those of the lower-order compositions.

The present effort to identify interesting three-cation systems and compositions has an understood possibility for false-negative detection of nontrivial three-cation phase behavior, e.g., when a three-cation phase has optical properties that are well modeled by linear interpolation from those of the composition subspaces. The workflow of Fig. 1 is designed to identify the most optically interesting phases, and the extent by which the data science models can guide materials discovery efforts depends on how closely the chemical physics underlying the target properties is related to that of optical absorption. For example, guidance of experimentation for discovery of novel electronic materials may be better served by adapting the methods of the present work to infrared measurements to characterize free carrier absorption or by moving beyond optical measurements by employing impedance spectroscopy or other measurements for which high-throughput methods have been developed (50). With regard to false-positive detection for discovery of three-cation phases, unambiguous demonstration of a false positive requires demonstration that the third cation is not present in an observed one or two-cation oxide phase, which is very difficult using the inkjet-printed samples. The primary expected cause of a false positive would be due to changes in optical absorption that arise from extrinsic properties such as film morphology rather than intrinsic optical properties of the metal oxide composition. The non-negative matrix factorization (NMF) model for extracting optical signals was designed to mitigate this issue, but ultimately, the tolerance for false positives must be considered in the context of a specific research goal. For example, to lower the false-negative rate at the expense of a higher false-positive rate, the down-selected compositions of interest can be expanded by considering candidate phase diagrams from several different values of K and/or using a higher threshold for $\log P$.

The identification of interesting composition-optical property trends in three-cation composition spaces can serve as a prior or a down-selection criterion for synthesis and/or measurement of properties that are more resource intensive than those employed in this high-throughput workflow. For example, choosing a threshold $\log_{10} P$ of -30 (see Fig. 2E) would down-select the $\sim 8\%$ of the three-cation composition spaces that are the most promising for discovery of exceptional properties. When applying this threshold to the individual three-cation compositions from the 10 at.% discretization, only 1% of compositions meet this criteria, providing 100-fold down-selection of compositions that merit further characterization. With this approach, one does not need to make automated high-throughput methods for measuring every material property of interest. Harnessing the interrelationships among properties through their shared relationships to composition and structure enables the high-throughput measurements of one property to guide exploration of other properties.

Materials and Methods

Printing, Calcination, and Imaging of Material Libraries. Methods described previously were adapted for inkjet printing of three-cation oxide libraries (46). For the present work, extensive preliminary experimentation was performed to minimize the printed feature size while maximizing the reproducibility of printer performance. The composition lines were deposited at the maximum print resolution of the printer (1,440 \times 2,880 dpi) using a CMYK tiff image at 300 dpi. The gradients were discretized into strips of

0.25 \times 0.25 mm (3 \times 3 pixels) samples of constant composition. The color was set to 50% of saturation to reduce the amount of ink delivered to produce sharp printed features, as delivery of larger quantities of ink produced wider strips than designed by spreading of the wet ink. The color saturation in the printing image is specified as an integer from 0 to 127 for each print channel. For each pixel, the 127 units of ink were distributed among the C, M, and Y channels to correspond to the desired composition. For two-cation composition lines, this corresponds to an integer change of 1 (circa 0.79 at.%) between neighboring samples and an average composition gradient of 1/127/0.25 mm, which is ~ 3.1 at.%/mm. For three-cation composition lines, this same composition gradient was applied, and the integer value of each channel was chosen by the set of 3 integers that sum to 127 and minimize the rms distance to the desired normalized composition. The printed composition lines were separated by a 0.5-mm region (6 pixels) with no ink deposition.

The inks were prepared from an ink base consisting of 12 g or Pluronic F127 dissolved in 500 mL of absolute ethanol, 16.0 mL of glacial acetic acid, and 6.0 mL of concentrated nitric acid (HNO₃) (46). The metal oxide precursor inks were prepared immediately before each print session by dissolving 3.3 mmol of each metal salt (used as received from Sigma-Aldrich) in 20.0 mL of the prepared ink base. The specific metal precursors were Fe(NO₃)₃·9H₂O, Mg(NO₃)₂·6H₂O, Ni(NO₃)₂·6H₂O, SnCl₂·2H₂O, TaCl₅, Y(NO₃)₃·6H₂O, Cu(NO₃)₂·3H₂O, InCl₃·4H₂O, Ce(NO₃)₃·6H₂O, and Co(NO₃)₂·6H₂O. Each three-cation composition library was printed onto a 10 cm \times 15 cm glass plate Corning Eagle XG aluminosilicate glass by assigning the different ink channels to the CMY colors of the library design image as needed. Using these inks with the printer parameters described, 1.9 nmol of metal oxide precursor are deposited per square millimeter within the printed strips. For the oxide products, an average thickness of ~ 10 nm is calculated from the bulk densities of the metal oxides; however, the film thickness variations inherent in the printing method produce regions of the film up to 50 nm in thickness. Thus, the films are thin enough to achieve equilibrium with the O₂(g) atmosphere during calcination at 750 °C for 10 h, which are the thermal processing conditions used herein.

A given print session corresponds to selection of six of these inks to load into six of the eight printer channels. Given the choice to explore three-cation spaces and the experiment design of depositing 10-fold duplication of the compositions from a three-cation space on a single library plate, the choice of using six elements in a single print session is motivated by the ability to accommodate batches of 20 library plates (1 for each of the 6-choose-3 elemental combinations) in each experiment process. Furthermore, the eight-channel printhead typically fails one channel at a time, and the planned use of six channels in a given print session extends the useful lifetime of the print head. For some print sessions, a subset of the 20 possible three-cation combinations were deposited.

The printed plates from a given print session were loaded into quartz racks placed in a 37 °C oven as each was printed. Each plate was supported on a quartz (fused silica) shelf to provide mechanical support and prevent redeposition of evaporated material onto the back side of the plates. The 20 three-metal oxide composition plates were loaded into three racks and held at 37 °C for 16 to 24 h, moved to a 67 °C oven for 24 to 36 h, and then calcined in a tube furnace as one batch. The calcination process was performed in a sealed 32 L tube using a controlled ramp (1.17 °C/min) to 750 °C, held for 10 h, and then actively (but uncontrolled) cooled to <100 °C over several hours. The process was conducted with the samples in the furnace hot zone, surrounded by baffles, the pressure of O₂ added to the sealed tube at room temperature was increased from 300 Torr to 600 Torr at the end of the ramp, when the O₂ was replaced via evacuation and back-filling with 300 Torr of fresh O₂ at the start of the 750 °C soak. The high-throughput workflow described herein was exercised with calcination temperatures spanning 350 °C to 750 °C during workflow development, although the primary results of the present work use only the 750 °C condition to mitigate limitations in phase formation related to elemental diffusion.

After calcination, each plate was imaged using an Epson V700 flat plate scanner in transmission positive film (slide film) mode at 4,800 dpi resolution. These plate images were examined using image processing software to increase contrast and magnification to verify the complete printing of the plate, without variation in ink delivery as a result of printer or print head clogging or air bubbles. After passing this visual quality control step, the plates were imaged using the spectral microscope system.

Optical absorption was characterized by a purpose-built microscope consisting of a microscope base (Olympus BX53M), sCMOS image sensor (Andor Zyla 5.5), 2.5 \times 0.08 numerical aperture objective lens (Olympus 1-U2M921), motorized XY stage (Prior H105), automated acquisition software (Molecular Devices MetaMorph), and a custom high-speed nine-channel hyperspectral light source (Advanced Research Consulting). The light source featured nine individual light emitting diodes (385 nm, 395 nm, 450 nm, 530

nm, 590 nm, 615 nm, 660 nm, 740 nm, 850 nm, with spectral half width from 10 to 40 nm) coupled via polychroic mirrors to a single liquid light guide connected to the microscope.

The high-throughput synthesis workflow was designed so that each library plate and their composition samples would experience the same conditions so that compositional variations can be modeled with all else being equal. This is most well achieved within a given print session, whereas additional "hidden" variables between different print sessions can introduce universal and/or composition-dependent variations in the synthesized materials. Possibilities for such variables include variation in precursor chemical lot, print-head characteristics, ambient humidity, and variable base pressure of the tube furnace that influences trace gas concentrations. For the data models of the present work, some variability in materials among print sessions is tolerable, although we note that adaption of the workflow to measure, for example, quantitative absorption coefficients for all three-cation oxides would require additional calibrations for a given print session and cross-print-session validation.

Image Preprocessing. Plates were imaged in a 17×26 grid of partially overlapping, nine-channel, 16-bit images. The raw, unstitched images comprised 43 GB of image data per plate. These images contain transmission intensity $I_{T,j}$ for which zero-valued pixels are removed by replacing intensity values below 10^{-9} with this minimal value. Since fewer than 20% of the pixels are designed to have printed material, the representative transmission signal from the bare glass plate, $I_{T, \text{glass}}$, is taken as the 90th percentile of the intensities in the set of images I_T . The dark signal from the detector is acquired by performing a scan with no illumination, producing the average image I_{dark} . The fractional transmission images are then calculated as $T_j = (I_{T,j} - I_{\text{dark}}) / (I_{T, \text{glass}} - I_{\text{dark}})$. The series of transmission images are stitched together to create the full plate image of fraction transmission, T . Joining this image to the plate design is achieved with an affine transformation calculated from manual alignment of three corners of the approximately rectangular printed region. This alignment procedure is required since the printer can exhibit variation in its print resolution (pixels per unit length) on the scale of 1% from day to day, which constitutes a negligible variation in the intended composition gradient but an impactful alteration to the association of imaged material to intended composition. We discretize the pseudocontinuous composition lines into small regions with approximately constant composition, which are referred to as "samples."

Calculating the Absorption Coefficient. The Beer-Lambert law provides a common model for calculating the spectral absorption coefficient α from fractional transmission and reflection:

$$\alpha \tau \cong -\ln(T),$$

where τ is the thickness of the material, and T is the fraction of incident light that is transmitted, and the amount of reflected light is assumed to be small compared to absorbed or transmitted light.

The experiment design (SI Appendix) includes printing of at least 10 samples of each composition at different locations on the plate. When calculating the absorption coefficient, we begin by selecting a set of pixels to be used in the absorption coefficient calculation. When calculating the absorption spectrum for each sample, this set of pixels ($\mathbb{P}_{\text{sample}}$) is the set of 3 \times 3 pixels associated with a given sample. The resulting sample-level absorption spectrum (α) is used in the phase diagram model, where the reconstruction loss is calculated on a per-sample basis.

When calculating a single absorption spectrum to represent each unique composition, the aggregation over the printed composition duplicates is achieved by collectively considering the set of pixels (\mathbb{P}_{comp}) in all samples with the respective composition. This composition-aggregated absorption coefficient (α_{comp}) was used both to generate ternary composition images and for the emergent property model.

Beginning with a set of pixels \mathbb{P} (either $\mathbb{P}_{\text{sample}}$ or \mathbb{P}_{comp} , as described previously), we take the negative logarithm of each fraction transmission value. We represent the resultant image as an (m channel \times n pixel) non-negative matrix Y . We find the best approximation to $Y \cong \hat{Y} = \alpha \cdot \tau$, where α is a nonnegative ($m \times 1$) matrix representing the absorption spectra and τ is a nonnegative ($1 \times n$) matrix representing the thickness of material deposited in a given pixel. To find α and τ , we use nonnegative matrix factorization with a loss function given by $\frac{1}{2} \|\hat{Y} - Y\| + a|\alpha| + b|\tau|$, where $\|\cdot\|$ denotes the Frobenius norm and $|\cdot|$ denotes the L1 norm. Heuristic choices of the regularization parameters a and b were $1e-6$ and $1e-2$, respectively. The result of this matrix factorization is a thickness profile τ with arbitrary units, giving the spectral absorption α units of inverse thickness. The thickness units from this model depend on the regularization parameters, and the use

of constant parameters enables analysis of the composition-dependent α while disregarding τ for the present purposes.

We present two methods for identifying optically interesting materials that are candidate three-cation phases, including substitutional alloys.

The Phase Diagram Model. A phase diagram can be described via a set of phases $\mathbf{p} = \{p_j\}$, where a phase $p_i = (c_i, e_i)$ consists of its composition c_i and its associated energy e_i . The set of phases that are thermodynamically stable are given by the convex hull of this space, whose calculation via the Quickhull algorithm (51) provides the phase diagram, where the facets of the convex hull are phase fields.

For the pseudoternary phase diagrams in this paper, we represent the composition c of a material as a 3-vector of molar fractions that sum to 1. The set of phases must include the three elemental phases $\mathbf{p}_{\text{elem}} = \{p_{\text{elem}_1}, p_{\text{elem}_2}, p_{\text{elem}_3}\}$ at each corner, to which we assign a constant energy $e_{\text{elem}_1} = e_{\text{elem}_2} = e_{\text{elem}_3} = 0$ so that the energy surface is constant through the composition graph in the absence of an additional phase. Constraining the energies e_{nonelem} of any additional nonelemental phases $\mathbf{p}_{\text{nonelem}}$ to satisfy $e_{\text{nonelem}_j} < 0$ ensures that all elemental phases will appear on the convex hull.

We used the following process to generate a set of candidate phase diagrams. We divided ternary composition space into a grid with composition intervals of 1/6, yielding a total of 25 potential nonelemental compositions (SI Appendix, Fig. S11A). We enumerated all sets of 25 choose n compositions with $n \leq 5$. For each set of compositions $c_{\text{non_elem}}$, we paired each nonelemental composition $c_{\text{non_elem}_j}$ with an energy $e_{\text{non_elem}_j}$ drawn uniformly at random from $(-1, 0)$ to create sets of nonelemental phases $\mathbf{p}_{\text{non_elem}}$. We appended the set of elemental phases and computed a phase diagram. For each set of 25 choose n nonelemental compositions, we repeated the random sampling of energies 100 times. This procedure yielded a total of 2,590,093 phase diagrams, although the number of unique phase diagrams is considerably smaller and depends on the random energy sampling. We note that the energies $\{e_j\}$ are not intended to be estimations of relative formation energies of the phases but rather provide a mechanism for random sampling of the relative formation energy, which results in random sampling of the set of Alkemade lines and thus compatibility triangles.

Given a phase diagram with phases $\mathbf{p} = \{p_j\}$, we make the assumption that the absorption spectra within each phase field is given by linear interpolation (i.e., the lever rule) of the absorption spectra of each phase $\alpha(p_j)$. For a given set of $\alpha(p_j)$, this assumption allows us to construct a linear model for the absorption spectrum at every composition α_{fit} . Fitting a phase diagram to observed data is a matter of selecting the values $\alpha(p_j)$ that minimize a given loss function. For each three-cation system, we observed a 9 channel \times 10 duplicate \times 3,306 composition absorption spectra α . We considered a loss function based on direct reconstruction of the signal with an \mathcal{L}_2 norm:

$$\text{loss}_{\text{signal}} = \sum_{\text{channels, duplicates, compositions}} (\alpha_{\text{fit}} - \alpha)^2.$$

This so-called "signal fit" has the disadvantage that 1) the lower photon energies result in smaller absorption signals, thus contributing less to the loss, and are underweighted in the fit, and 2) a composition with high variability in signal among 10 composition duplicates should be devalued to avoid bias toward less trustworthy data in the optimization. We address these issues by computing the SD of the signal $\sigma(c)$ over nearby compositions: Representing three-cation compositions as points on an equilateral triangle where each vertex is a pure element, we normalize distances between the vertices to a distance of 1. We compute $\sigma(c)$, the 9-channel SD of the absorption spectra at a given composition c , considering each α measurement whose composition lies within a radius of 0.05 of c , including all duplicates. This enables regression fit of α_i to minimize the loss:

$$\text{loss}_{\text{sigma}} = \sum_{\text{channels, duplicates, compositions}} \frac{(\alpha_{\text{fit}} - \alpha)^2}{\sigma(c)}.$$

This so-called "sigma fit" was used for all results in the manuscript, and signal fit results are also provided in the dataset release. We fit all the candidate phase diagrams, and for each N in $N = 4, 5, \dots, 8$, we saved the fits with the smallest loss.

The Emergent Property Model. The purpose of this model is to identify three-cation compositions whose optical absorption spectrum is difficult to attribute to a linear combination of one or two-cation oxides.

This section uses absorption spectra α_{comp} obtained by combining all 10 experimental copies of a composition on a plate prior to nonnegative matrix factorization. We discretize the ternary composition space into composition intervals of 1/10, corresponding to 30 regions on the perimeter $r_{\text{perimeter}}$ and 36 internal regions r_{interior} (SI Appendix, Fig. S11B).

For each perimeter composition, we define the absorption spectrum of the neighboring region $r_{\text{perimeter}}$ as the mean of the absorption spectra measurements for compositions within $r = 1/10$ of the grid composition,

$$\alpha_{r_{\text{perimeter}}} = \frac{\sum_{\text{CEF}} \alpha}{n_{r_{\text{perimeter}}}}$$

where $n_{r_{\text{perimeter}}}$ is the number of α measurements in $r_{\text{perimeter}}$.

We enumerate an exhaustive list of linear combinations of three or fewer absorption spectra of these perimeter regions. For all (30 choose 3) combinations of perimeter regions, we enumerate all possible linear combinations of the absorption spectra in a discretized ternary composition space with intervals of 1/10. This results in a set of 150,105 absorption spectra that represent all possible absorption spectra, α_{comb} , one could obtain by combining up to three of the one and two-cation compositions. For any material whose absorption spectrum is distinct from every element of this set, we can assert that the material does not comprise a simple mixture of the one and two-cation phases.

For each interior region r , we compute the nine-channel mean α_r and the nine-channel SD σ_r of the absorption spectra of the compositions in the region. Since the samples in each region correspond to compositions within a composition window of ~ 10 at.%, the SD signal contains a substantial contribution from systematic variation in α only when the compositional gradient of α is particularly high. The more consistently significant contributions to the SD signal arise from sample-to-sample variability in the signal and any variation in signal from compositions within this window that belong to different composition-gradient lines in the experiment design (SI Appendix, Figs. S1C and S11B). For each $\alpha \in \alpha_{\text{comb}}$, we compute the probability $p(\alpha|N(\alpha_r, \sigma_r))$ that α was drawn from a nine-dimensional Gaussian distribution centered at α_r with SD σ_r . The emergent property score of the interior region is the maximum $p(\alpha|N(\alpha_r, \sigma_r))$ over all $\alpha \in \alpha_{\text{comb}}$.

Sputter Deposition and Characterization. The Co-Ta-Sn follow-up measurements in Fig. 4 are based on a continuous composition spread synthesized atop a 100-mm-diameter XG glass disk by reactive cosputtering of metal targets (Co, Ta, and Sn) using radio-frequency (RF) power supplies in a combinatorial sputtering system (Kurt J. Lesker, CMS24). The sputtering atmosphere was composed of 0.6 mTorr reactive O_2 gas and 5.4 mTorr inert Ar sputtering gas. The RF powers were adjusted to obtain the designed composition in the wafer center, and the nonconfocal geometry of sputter sources provided a continuous composition gradient spanning a 60 to 70 at.% range in the concentration of each cation element across the XG glass substrates. Postdeposition annealing proceeded in a tube furnace at 750 °C for 10 h (the same protocol as the printed library plate).

The library compositions were obtained by XRF measurements using an EDAX Orbis Micro-XRF system with an X-ray beam ~ 2 mm in diameter with metal loadings ($\text{nmol}\cdot\text{cm}^{-2}$) based on elemental calibrations using commercial XRF standards (Micromatter). For each composition, the cation-weighted molar density was calculated using the molar density of the endmember phases (Co_2O_3 , Ta_2O_5 , SnO_2). This average molar density was then combined with the cation molar loading to estimate the film thickness,

which ranged from 330 to 520 nm. Each calculated thickness has $\sim 10\%$ uncertainty, which is negligible compared to the dynamic range of absorption coefficient values observed in Fig. 4C.

The XRD was performed using a Bruker DISCOVER D8 diffractometer with $\text{Cu K}\alpha$ radiation from a Bruker I μ S source. The measurements used a 0.3-mm collimator to acquire a diffraction signal on a sample region of about 0.3 mm \times 2 mm with two-dimensional VANTEC-500 detector followed by integration into one-dimensional patterns using DIFFRAC.SUITE EVA software. The crystalline phases present in each sample are identified by matching XRD patterns with entries in the International Crystallography Diffraction Database in the EVA software.

Gaussian function fitting was used on one-dimensional XRD patterns to obtain the peak position (2θ) for rutile (101) reflection $2\theta_{101}$ between 32° and 36° and (211) reflection $2\theta_{211}$ between 41° and 54°. Next, we solved the lattice parameters a and c as follows:

$$\begin{cases} 2d_{101} \sin \theta_{101} = \lambda \\ 2d_{211} \sin \theta_{211} = \lambda \end{cases} \Rightarrow \begin{cases} a = \frac{\lambda}{\sqrt{\sin^2 \theta_{211} - \sin^2 \theta_{101}}} \\ c = \frac{\lambda}{\sqrt{5 \sin^2 \theta_{101} - \sin^2 \theta_{211}}} \end{cases}$$

where λ denotes the X-ray wavelength (1.5406 Å).

Optical measurements on the sputtered film were performed using a Shimadzu Solidspec-3700 spectrophotometer. Transmittance (T) and diffuse reflectance (R) measurements were performed separately on each sample spot using the integrating sphere. A BaSO_4 powder reflector was used as the reflection standard. Together with the sample thickness (t_{XRF}) from XRF, the absorption coefficient was calculated as follows:

$$\alpha = \frac{1}{t_{\text{XRF}}} \ln \left(\frac{T}{1-R} \right)$$

Data Availability. The optical absorption spectra, fitted phase diagrams, and mixture probabilities have been deposited in Google Cloud Storage (<http://storage.googleapis.com/gresearch/metal-oxide-spectroscopy/README.txt>; see SI Appendix for documentation and access instructions).

ACKNOWLEDGMENTS. This material is based on work performed by the Joint Center for Artificial Photosynthesis, a Department of Energy Energy Innovation Hub, supported through the US Department of Energy Office of Basic Energy Sciences under Award DE-SC0004993, which supported materials synthesis and characterization experiments. Google Applied Science supported the development and execution of the computational workflow as well as procurement of the hyperspectral microscope. Structural characterization and follow-up validation experiments were supported by the US Department of Energy Office of Basic Energy Sciences under Award DE-SC0020383. We are grateful for helpful discussions and guidance in the development of the computational workflow from Muskaan Goyal, Eric Christiansen, Edward A. Baltz, Derek Leong, Austin Blanco, and Mike Ando (Google Applied Science). We also appreciate the support of the experiment workflow from Edwin Soedarmadji (Caltech). We additionally appreciate helpful suggestions by David Fork and Michael Brenner (Google Applied Science).

- B. L. DeCost *et al.*, Scientific AI in materials science: A path to a sustainable and scalable paradigm. *Mach. Learn. Sci. Technol.* **1**, 033001 (2020).
- J. E. Saal, A. O. Olynyk, B. Meredig, Machine learning in materials discovery: Confirmed predictions and their underlying approaches. *Annu. Rev. Mater. Res.* **50**, 49–69 (2020).
- T. Latinovic, D. Preradović, C. R. Barz, A. P. Vadean, M. Todić, Big Data as the basis for the innovative development strategy of the Industry 4.0. *IOP Conf. Ser.: Mater. Sci. Eng.* **477**, 012045 (2019).
- C. Sutton *et al.*, Identifying domains of applicability of machine learning models for materials science. *Nat. Commun.* **11**, 4428 (2020).
- O. A. von Lilienfeld, K. Burke, Retrospective on a decade of machine learning for chemical discovery. *Nat. Commun.* **11**, 4895 (2020).
- Y. Yao *et al.*, High-throughput, combinatorial synthesis of multimetallic nanoclusters. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 6316–6322 (2020).
- Y. Kim, E. Kim, E. Antonio, B. Meredig, J. Ling, Machine-learned metrics for predicting the likelihood of success in materials discovery. *NPJ Comput. Mater.* **6**, 131 (2020).
- S. D. Conradson *et al.*, Nonadiabatic coupling of the dynamical structure to the superconductivity in $\text{YSr}_2\text{Cu}_2.75\text{Mo}_{0.25}\text{O}_{7.54}$ and $\text{Sr}_2\text{CuO}_{3.3}$. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 33099–33106 (2020).
- V. I. Hegde, M. Aykol, S. Kirklın, C. Wolverton, The phase stability network of all inorganic materials. *Sci. Adv.* **6**, eaay5606 (2020).
- H. Zhang *et al.*, Application of fuzzy learning in the research of binary alloys: Revisit and validation. *Comput. Mater. Sci.* **172**, 109350 (2020).
- C. Chen, Y. Zuo, W. Ye, X. Li, S. P. Ong, Learning properties of ordered and disordered materials from multi-fidelity data. *Nat. Comput. Sci.* **1**, 46–53 (2021).
- J. Noh, G. H. Gu, S. Kim, Y. Jung, Machine-enabled inverse design of inorganic solid materials: Promises and challenges. *Chem. Sci. (Camb.)* **11**, 4871–4881 (2020).
- A. Zunger, Inverse design in search of materials with target functionalities. *Nat. Rev. Chem.* **2**, 0121 (2018).
- L. Yang *et al.*, Hyperspectral microscopy of complex metal oxides. Metal-oxide-spectroscopy. https://www.google.com/search?q=gs%3A%2F%2Fgsearch%2Fmetal-oxide-spectroscopy&rlz=1C1NHXL_enUS832US832&oq=gs%3A%2F%2Fgsearch%2Fmetal-oxide-spectroscopy&aqs=chrome.0.69i59j69i58.943j0j7&sourceid=chrome&ie=UTF-8. Deposited 25 August 2021.
- S. P. Ong *et al.*, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
- NIST Inorganic Crystal Structure Database, NIST Standard Reference Database 3. National Institute of Standards and Technology. <https://icsd.products.fiz-karlsruhe.de/>. Accessed 15 March 2021.
- J. C. Védrine, Heterogeneous catalysis on metal oxides. *Catalysts* **7**, 341 (2017).
- F. F. Abdi, S. P. Berglund, Recent developments in complex metal oxide photoelectrodes. *J. Phys. D Appl. Phys.* **50**, 193002 (2017).

19. L. Zhou *et al.*, Successes and opportunities for discovery of metal oxide photoanodes for solar fuels generators. *ACS Energy Lett.* **5**, 1413–1421 (2020).
20. G. Cheon *et al.*, Revealing the spectrum of unknown layered materials with super-human predictive abilities. *J. Phys. Chem. Lett.* **9**, 6967–6972 (2018).
21. W. Li, R. Jacobs, D. Morgan, Predicting the thermodynamic stability of perovskite oxides using machine learning models. *Comput. Mater. Sci.* **150**, 454–463 (2018).
22. G. Hautier, C. C. Fischer, A. Jain, T. Mueller, G. Ceder, Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chem. Mater.* **22**, 3762–3767 (2010).
23. K. Alberi *et al.*, The 2019 materials by design roadmap. *J. Phys. D Appl. Phys.* **52**, 013001 (2019).
24. E. O. Gomes *et al.*, Computational procedure to an accurate DFT simulation to solid state systems. *Comput. Mater. Sci.* **170**, 109176 (2019).
25. C. Nyshadham, *et al.*, Machine-learned multi-system surrogate models for materials prediction. *NPJ Comput. Mater.* **5**, 51 (2019).
26. B. Nestler, H. Garcke, B. Stinner, Multicomponent alloy solidification: Phase-field modeling and simulations. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **71**, 041609 (2005).
27. D. Lerch, O. Wieckhorst, G. L. W. Hart, R. W. Forcade, S. Müller, UNCLE: A code for constructing cluster expansions for arbitrary lattices with minimal user-input. *Model. Simul. Mater. Sci. Eng.* **17**, 055003 (2009).
28. A. W. Sleight, Chemistry of high-temperature superconductors. *Science* **242**, 1519–1527 (1988).
29. P. A. Wright, *Microporous Framework Solids* (Royal Society of Chemistry, 2008).
30. D. R. Sutherland *et al.*, Optical identification of materials transformations in oxide thin films. *ACS Comb. Sci.* **22**, 887–894 (2020).
31. X. Liu *et al.*, Inkjet printing assisted synthesis of multicomponent mesoporous metal oxides for ultrafast catalyst exploration. *Nano Lett.* **12**, 5733–5739 (2012).
32. A. R. Oganov, C. J. Pickard, Q. Zhu, R. J. Needs, Structure prediction drives materials discovery. *Nat. Rev. Mater.* **4**, 331–348 (2019).
33. J. Schmidt *et al.*, Predicting the thermodynamic stability of solids combining density functional theory and machine learning. *Chem. Mater.* **29**, 5090–5103 (2017).
34. E. Reddington *et al.*, Combinatorial electrochemistry: A highly parallel, optical screening method for discovery of better electrocatalysts. *Science* **280**, 1735–1737 (1998).
35. M. Woodhouse, G. S. Herman, B. A. Parkinson, Combinatorial approach to identification of catalysts for the photoelectrolysis of water. *Chem. Mater.* **17**, 4318–4324 (2005).
36. D. Seley, K. Ayers, B. A. Parkinson, Combinatorial search for improved metal oxide oxygen evolution electrocatalysts in acidic electrolytes. *ACS Comb. Sci.* **15**, 82–89 (2013).
37. J. N. Cawse, Experimental strategies for combinatorial and high-throughput materials development. *Acc. Chem. Res.* **34**, 213–221 (2001).
38. P. J. McGinn, Combinatorial electrochemistry—Processing and characterization for materials discovery. *Mater. Discov.* **1**, 38–53 (2015).
39. R. A. Potyrailo, V. M. Mirsky, Combinatorial and high-throughput development of sensing materials: The first 10 years. *Chem. Rev.* **108**, 770–813 (2008).
40. W. F. Maier, K. Stöwe, S. Sieg, Combinatorial and high-throughput materials science. *Angew. Chem. Int. Ed. Engl.* **46**, 6016–6067 (2007).
41. R. Potyrailo *et al.*, Combinatorial and high-throughput screening of materials libraries: Review of state of the art. *ACS Comb. Sci.* **13**, 579–633 (2011).
42. A. Ludwig, Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods. *NPJ Comput. Mater.* **5**, 70 (2019).
43. M. L. Green, I. Takeuchi, J. R. Hattrick-Simpers, Applications of high throughput (combinatorial) methodologies to electronic, magnetic, optical, and energy-related materials. *J. Appl. Phys.* **113**, 231101 (2013).
44. J.-C. Zhao, Combinatorial approaches as effective tools in the study of phase diagrams and composition–structure–property relationships. *Prog. Mater. Sci.* **51**, 557–631 (2006).
45. X.-D. Xiang, Combinatorial materials synthesis and high-throughput screening: An integrated materials chip approach to mapping phase diagrams and discovery and optimization of functional materials. *Biotechnol. Bioeng.* **61**, 227–241 (1998-1999).
46. J. A. Haber *et al.*, Discovering Ce-rich oxygen evolution catalysts, from high throughput screening to water electrolysis. *Energy Environ. Sci.* **7**, 682–688 (2014).
47. A. Shinde *et al.*, Discovery of Fe-Ce Oxide/BiVO₄ photoanodes through combinatorial exploration of Ni-Fe-Co-Ce oxide coatings. *ACS Appl. Mater. Interfaces* **8**, 23696–23705 (2016).
48. Y. Hu *et al.*, First principles calculations of intrinsic mobilities in tin-based oxide semiconductors SnO, SnO₂, and Ta₂SnO₆. *J. Appl. Phys.* **126**, 185701 (2019).
49. L. Zhou *et al.*, Rutile alloys in the Mn–Sb–O system stabilize Mn³⁺ to enable oxygen evolution in strong acid. *ACS Catal.* **8**, 10938–10948 (2018).
50. R. B. van Dover, L. F. Schneemeyer, R. M. Fleming, Discovery of a useful thin-film dielectric using a composition-spread approach. *Nature* **392**, 162–164 (1998).
51. C. B. Barber, D. P. Dobkin, H. Huhdanpaa, The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* **22**, 469–483 (1996).