Supplementary Information

Discovery of complex oxides via automated experiments and data science

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

\textsuperscript{a}Google Applied Science, Mountain View, CA. \textsuperscript{b}Division of Engineering and Applied Science and Joint Center for Artificial Photosynthesis, California Institute of Technology, Pasadena, CA 91125.

\textsuperscript{1}To whom correspondence may be addressed. Email: gregoire@caltech.edu

A. Design of high throughput synthesis and generality for high-order composition spaces

While the present work focuses on mapping 3-cation composition spaces, the design of experiments is generalizable to any number of cations. The geometrical construct of phase fields is such that for \( N \) components the traditional composition graph is drawn in \((N-1)\) dimensional space. The phase field boundaries are \((N-2)\) dimensional objects in the space, e.g., for 3-cation phase diagrams the phase boundaries are lines. Composition lines \((1\)-dimensional) spanning the composition space will generally intersect the phase field boundaries at single composition points \((0\)-dimensional). These geometric constructs are illustrated for \( N=3 \) in Figure S1a and \( N=4 \) in Figure S1b. The dimensionality of the composition lines and their intersection with phase field boundaries does not depend on \( N \), making the synthesis of linear composition gradients spanning the composition space an effective strategy for identifying the existence of phase boundaries within the \( N \)-component composition space.

Since phase field boundaries can generally exist at any position and angle within the composition graph, synthesis of multiple composition lines spanning the space provides ample opportunities for detection of phase boundaries. For the study of 3-component spaces, we designed the set of composition lines shown in Figure S1c. We first choose the 7 equally-space points along each edge of the composition graph, for example the composition \( \text{A}_{1-x}\text{B}_x \) where \( x=0,1,2,3,4,5,6 \). Then, a composition line is drawn from each composition with even numbered \( x \) to each composition on the other 2 boundaries with odd numbered \( x \). For each of the 3 end member \((1\)-component vertex, \( x = 0 \) or 6) compositions there are 3 composition line segments, and for each of the 6 2-component compositions with even numbered \( x \) \((x=2, 4)\) there are 6 composition line segments. The set of 45 unique composition line segments that span the ternary composition space are shown in Figure S1c, and combined with the 2-component edges of the composition graph, provide the 48 composition line segments synthesized for each composition library.

The inkjet printing strategy is based on the dual considerations of (i) printing composition lines with gradual composition gradients and (ii) incorporating 10 duplicates of each of the 48 composition line segments on a single \( 10\times15 \) cm glass plate. These considerations are well met by printing lines that each contain a series of connected composition line segments from the set of 48. For example, Figure S2a shows the 10 unique composition paths that originate from one end-member composition. The full set of printed lines on the composition plate are constructed by applying this set of composition paths to the other two end member compositions, constructing similar paths originating from the 2-component compositions noted above with even numbered \( x \), and duplicating these composition paths with alternating directions (red-to-blue and blue-to-red in the illustration of Figure S2). In total, 84 printed composition lines are required to provide the 10
duplicates of each of the composition line segments, and randomizing the order of these lines provides the print design shown in Figure S2c where the color of each pixel corresponds to the composition using the A=cyan, B=yellow, C=magenta color scheme. The synthesis then proceeds by loading this image in the printer and assigning the color channels to the desired elemental inks.

While the present work’s focus on 3-component systems enables comprehensive consideration of all model phase diagrams (up to the discretization of the composition space), we note that the experiment workflow is scalable to any number of components. Since the combinatorics of higher-order spaces can impose considerable computational expense to perform the iterations over all possible phase diagrams, scaling the current methods may require utilizing the analysis scheme envisioned in Figure S1 in which composition gradient lines enable detection of phase field boundaries. To demonstrate this concept, Figure S3 shows the optical data for the Fe-In-Y oxide system, along with a model phase diagram. For the chosen composition line segment, the phase diagram model indicates 3 crossings of phase field boundaries. The assumption of linear variation in molar concentration of phases within phase fields combined with assumption that $\alpha$ for a given composition is the molar-weighted sum of the respective values for the constituent phases, any traversal of a phase boundary should produce a change in slope of the $\alpha$ vs composition plot. Furthermore the change in the slope will be proportional to the sine of the angle between the phase boundary line and the composition gradient line. In Figure S3, this is evident by the relatively minor change in signal near the middle phase boundary transition point compared to the other two phase boundary transitions occurring at larger angles. In general, this demonstrates how designing composition lines to span the composition space in different composition directions can facilitate discovery of phase boundaries.

B. Visualization for quality control

Identification of various experimental artifacts was aided by difference-from-median plots generated by subtracting from each sample’s $\alpha$ the median value over the 10 composition duplicates. The example image shown in the workflow of Figure 1 (step 8) is indicative of sporadic printing, probably an intermittently operational printhead, providing an example of halting the pipeline based on poor quality data. Other representative examples of these plots are shown in Figure S4. Given the particulate morphology of the resulting films, some variation is expected and is precisely the motivation for performing phase analysis on the aggregation of at least 10-fold duplication of every composition. As a result, Figure S4a is an example of a print with now observable artifacts, only the expected deviations from the median signals that are quasi-random with respect to position. Figure S4b shows an artifact from a crack in the glass substrate, which probably doesn’t strongly influence the aggregate signal but will artificially increase the assessed variation in signal for the compositions that fall along the crack, which is sufficient cause to classify this plate as invalid. Figure S4c shows a systematic trend where the lower-left and upper-right exhibit higher and lower absorption, respectively. This could be from an unintended temperature or gas flow gradient during annealing or a non-level plate during printing that causes more diffuse, thinner deposits in the upper-right and more focused, thicker deposits in the bottom-left. Given the composition duplications and random ordering of the printed rows, this gradual systematic shift has negligible influence on the composition trends of the optical signals aggregated over composition duplicates, which are the refined data that are critical to the analysis workflow. Figure S4d shows the composition reconstruction of the absorption signals from this plate, revealing smooth composition trends suitable for phase diagram modelling and probability calculations.
C. The thermodynamic equilibrium assumption and implications of substitutional alloying in the phase diagram model

The thermodynamic equilibrium assumption is based on the following reasoning. Gibbs’ phase rule for relating the number of components C and the number of thermodynamic degrees of freedom F to the number of phases P is

\[ F = C - P + 2. \]

For a 3-cation composition with oxygen there are 4 components (C). Concerning the number of phases, the equilibration at the annealing condition combined with the constant partial pressure of \( O_2 \) corresponds to gaseous \( O_2 \) being one phase that equilibrates with \( P_{\text{solid}} \) solid state phases. We assume that the temperature and pressure of the annealing condition do not happen to correspond to phase boundaries in the full composition, temperature, pressure phase diagram. This assumption means that for a given composition, temperature and pressure comprise 2 thermodynamic degrees of freedom (F) that can vary independently while remaining in the same (temperature, pressure, composition) phase field. In the presence of alloying, there may be 1 or 2 additional thermodynamic degrees of freedom, which we will call \( F_{\text{comp}} \).

Using \( C=4, F = 2 + F_{\text{comp}} \) and \( P = 1 + P_{\text{solid}} \), the equation becomes

\[ P_{\text{solid}} = 3 - F_{\text{comp}}. \]

Since \( F_{\text{comp}} \) is nonnegative, the maximum number of phases is 3, which is pertinent to both of the data science models. In the case of the Co-Ta-Sn rutile oxide alloy, the 3-cation alloy composition region has 2 degrees of freedom, where \( F_{\text{comp}} = 2 \) corresponds to the observed \( P_{\text{solid}} = 1 \). This alloy phase is also observed as a 2-phase mixture with either \( \text{Co}_3\text{O}_4 \) or \( \text{Ta}_2\text{O}_5 \), in which case \( F_{\text{comp}} = 1 \) means one cannot independently vary the composition on the cation sublattice, i.e. for a sublattice composition \( \text{Co}_x\text{Ta}_y\text{Sn}_{1-x-y} \), \( x \) and \( y \) are covariate in the 2-phase regions and not covariate in the 1-phase region. It is worth noting that if the annealing temperature happens to be the eutectoid temperature for a given composition, an additional phase could present (a total of 4 for a 3-cation composition), and the phase diagram model would not fully capture this particular situation.

As noted in the main text, we also assume that we do not melt our films, i.e. we avoid the liquidus portion of the phase diagrams and the alterations in film morphology that would likely result from freezing a melted metal oxide as opposed to crystallization within the solid state. Our choice of the 750 °C anneal temperature supports this assumption, because our prior characterization of the metal oxides synthesized from the individual precursors annealed at >550 °C revealed formation of the following phases, which are ordered from lowest to highest melting point: \( \text{Co}_3\text{O}_4, \text{CuO, Fe}_2\text{O}_3, \text{SnO}_2, \text{Ta}_2\text{O}_5, \text{In}_2\text{O}_3, \text{NiO, Ce}_2\text{O}_3, \text{Y}_2\text{O}_3, \text{MgO, Co}_3\text{O}_4 \) is the only one with melting temperature below 1300 °C. While its melting temperature of 895 °C does not preclude the existence of a eutectic at or below 750 °C, given the refractory cation elements being combined with Co, we assume that the fraction of compositions in our experiments that are molten at 750 °C is sufficiently small to have negligible impact on the data science models.

The above invocation of Gibbs’ phase rule models the annealing of each sample as an “open” system where lattice oxygen can be exchanged with \( O_2 \) in the annealing atmosphere. Either intentional control of the lattice oxygen concentration or a kinetic limitation, for example from limited diffusion of lattice oxygen, could result in a “closed” system in which the number of possible coexisting phases would increase by 1, which would substantially complicate the data science models as the composition graphs would need to be considered in the 4-component composition
space (including oxygen) where 3-cation compositions could have 4 coexisting phases instead of the maximum of 3 considered in the present work.

Substitutional alloying is not explicitly included in the phase diagram model because the set of possible phase diagrams is vastly larger than the set under the assumption of no alloying, as the latter involves tiling the composition space with triangles whereas the former lacks a similar geometric construct. To gauge the impact of this decision, we consider scenarios when alloying does and does not result in changes to the optical absorption properties. When the alloy composition has the same optical absorption properties as the stoichiometric phase, the error in the model phase diagram is that at the alloy composition (assumed to be near the stoichiometric composition, compared to the compositions of other single-phase regions) will be modeled as a mixture of the stoichiometric phase and other phases, whereas in reality there is the single alloy phase. The fitted model for the absorption spectrum will be the weighted sum of the phases, and the contributions to this sum from the other phases comprises the error in the model, which is negligible for alloying over small composition ranges, compared to composition spacing between phases and for small differences between the optical properties of the associated phases. If alloying does substantially alter optical absorption properties and the alteration is systematic with composition deviation from the stoichiometric phase, then the saturated alloy (the phase-pure composition most different from the stoichiometric phase) may be modeled as a separate phase. In this case, the compositions between the stoichiometric phase and the saturated alloy will be well modeled, and the errors will arise in neighboring mixed-phase regions where, for example, an alloy phase is mixed with a secondary phase but the phase diagram model includes a 3-phase mixture of the stoichiometric phase, saturated alloy, and secondary phase. To the extent that the alloy phase can be modeled as a mixture of the stoichiometric phase and the saturated alloy, the resulting error in the model is limited to deviations in the phase weights, which in the limit of small variation in optical properties over neighboring phases gives rise to modest errors in the model. As a result, the phase diagram models may be applied to datasets with substantial alloying with some anticipated differences between the measured and modeled absorption patterns in the alloying composition regions. Most importantly, some 2 and 3-cation phases in the resulting model phase diagram may not be entirely new structures but instead alloyed variants of other structures with substantially different optical properties.

D. X-ray diffraction characterization of combinatorial libraries

The diffraction signal from the thin metal oxide deposits is quite weak; thus, the diffraction signal is rarely sufficiently intense to enable unambiguous phase identification. The diffraction data shown here are emblematic of the strongest signals obtained from the combinatorial libraries, which limits the XRD validation of the model phase diagrams. The demonstration XRD experiments guided by analysis of optical data commences with the Ce-Fe-In oxide system, a system chosen during development of the workflow due to the known existence of a 3-cation phase, CuFeInO₄. The experimental and analysis workflow for this system differs from those of the primary results in this paper. Specifically, this composition library was annealed at 550 °C instead of 750 °C and the phase diagram modelling was performed with composition intervals of 1/7 instead of 1/6. Figure S5 summarizes the optical data, phase diagram model, and XRD results. The optical signals exhibit fairly monotonic composition trends at high photon energy, as evidenced by the composition map of $\alpha_{3.2eV}$. At lower photon energy, non-trivial trends within the 3-cation composition region indicate the presence of non-trivial 3-cation phase behavior with respect to optical properties, as evidenced by the composition map of $\alpha_{1.9eV}$. The $K=3$ model phase diagram is shown in Figure S5b, where the 6 phase points are labelled A through F. The 7 corresponding compositions, as well as 5 additional compositions were chosen for XRD characterization. These points and the corresponding primary and secondary (where applicable) phases are listed in Figure S5c. The Cu end-member phase (phase A) is expected to be CuO, although the diffraction signal is not sufficiently strong to identify any phase at this composition. With the addition of Fe, an Fe-alloyed variant of this structure is observed and denoted as phase $A^\ast$. Intuitively, Fe in this alloy should be nominally Fe²⁺ in a substantially different coordination environment compared to Fe³⁺ in phase D (CuFeO₄). As a result, one expects the (Cu,Fe)O alloy
to exhibit different optical properties than a mixture of CuO and CuFe$_2$O$_4$. While detailed inspection of this composition line shows some corresponding nonlinear variation in signal along this Cu-Fe composition line, the deviations of those signals from the linear model of the candidate phase diagram in Figure S5c were not sufficient to dedicate one of the $K = 3$ phase fit points to this alloy phase. Regarding the $K = 3$ phase diagram, this is false-negative phase detection, highlighting that the candidate phase diagram for a given $K$ is not intended to report all phases in the composition space, but rather the $K$ most important phases for explaining the optical absorption data.

The primary phase observed at each of the other 6 compositions labelled in Figure S5b is shown in Figure S5d. The XRD patterns and stick patterns for matched phases are shown in Figure S5e-S5g. The qualitative agreement between the predicted and observed phases and phase mixtures is excellent, with the most common inability to validate the prediction arising from the poor detectability limit, especially for minority phases. Phase E is identified to be the 3-cation phase CuFeInO$_4$ that was the initial synthesis target. It is observed at composition with less In than the formula unit, and the intensity of the XRD pattern is quite weak and only interpretable as the target phase due to the primary peaks being at scattering vector magnitudes distinguishable from those of the competing 1- and 2-cation phases. The phase point F shows a weaker version of this same pattern, which would require further characterization to fully understand, and for the present purposes is labelled as a Cu-rich variant of CuFeInO$_4$. If this phase does indeed support alloying into Cu-rich compositions, the optical characterization indicates that the alloying introduces trivial composition dependence of the optical absorption above 2.1 eV, and nontrivial dependence at lower photon energies.

E. Follow-up measurements on sputter-deposited Sn-Co-Ta oxide

The XRD results of Figure 4 are complemented by additional data on sputter-deposited Sn-Co-Ta in Figures S8-S10.

Characterization of oxygen evolution electrocatalysis proceeded in pH 3 and then pH 0 (Fig. S8) electrolyte, highly corrosive conditions where Co oxides are known to suffer from rapid corrosion. Using a combination of cyclic voltammetry and operation for 160 s at 1.7 V vs RHE, the catalytic activity was assessed. Complementary measurement of XRF before and after electrochemical testing indicated the extent of Co corrosion. The summary of these results (Fig. 4c) indicates that some low-Co-concentration samples exhibit no measurable catalytic activity, while the remainder exhibit a broad range of both catalytic activity and Co corrosion. Given the desirability for high current density and low corrosion, the nominal Pareto front for these parameters is shown in the figure, demonstrating how the 3-cation compositions push this front by exhibiting appreciable activity with substantial stabilization of the Co cations.

The visual transparency of Co-poor compositions in the sputter deposited library is supported by spectroscopy results in Fig. S8, the ensuing composition plots of optical absorption in Fig. 4c, and the ellipsometry characterization summarized in Figure S9. Variable-angle spectroscopic ellipsometry (VASE) measurements were performed using a J.A. Woollam V-VASE Ellipsometer using 3 thin film compositions from a duplicate of the library used for Fig. 4. Focusing probes were used for small spot measurements on the samples and translucent tape was applied to the back of the glass substrate (XG-glass) to minimize backside reflections. A reference measurement on pristine XG-glass substrate was also performed to obtain a reference Cauchy model for subsequent modeling of the experimental sample data. V.A.S.E. for Windows was used as the modeling software.

To extract the absorption coefficient (Figure S9a) from the ellipsometry data (Figure S9b-d) the sample was modeled as a stack of (1) XG-glass substrate / (2) sample layer / (3) surface roughness layer. For (1) the Cauchy Oscillator model of the reference substrate measurement was used. The model of the sample layer (2) consisted of a Cauchy oscillator (with enabled modeling of absorption for the visually transparent material (Sn$_0.56$Co$_{0.07}$Ta$_{0.35}$) or two Tauc-
Lorentz Oscillators for measurements of optically absorbing materials (Sn_{0.48}Co_{0.22}Ta_{0.30} and Sn_{0.42}Co_{0.32}Ta_{0.26}). To account for material inhomogeneities due to deposition process and pinholes/voids the sample layer was modeled in addition with simple index grading and with effective medium approximation between void material and sample oscillator model. The fit of the developed model to the sample data was performed in the photon energy regime (eV) between 1.0 eV and 4.0 eV. After successful fitting the absorption coefficient \( \alpha \) could be extracted from the sample layer (2). Figure S9a shows the extracted absorption coefficient of only the sample layer (2) with raw PSI data and corresponding model fit of complete layer stack.

F. Public Data Release
The optical absorption spectra, fitted phase diagrams, and mixture probabilities described in this paper are available at gs://greesearch/metal-oxide-spectroscopy. We released data from 6 print sessions, comprising 173 plates, 131 quaternary oxide systems, 6,918,024 individual composition samples, and 376,752 distinct compositions. While the ten-fold replicates within each plate are well controlled, uncontrolled variables (printhead age, etc) may lead to poorer consistency between print sessions.

The data exists in four directories and one metadata file. Each directory contains one type of data, with one *.csv file per printed plate.

i. The data in ten_replicas/ consists of optical transmission data, with one row per printed patch on a plate. The column headers are:
- ExplID: an integer experiment ID for the printed patch on the plate.
- row, col: The row and the column coordinates of the printed patch in the microscope image
- signal_: The measurement of \( \alpha \), the optical transmission spectrum of the printed patch, at a given wavelength. \# ranges from 0 to 8, inclusive, indicating transmission spectra at the following wavelengths: 375, 395, 455, 530, 590, 617, 660, 735, & 850 nm.
- plate: The integer plate identifier.
- line: An integer identifier of the composition gradient that was printed.
- line_experiment_id: An integer identifier of the composition sample along the composition gradient.
- replica: An integer identifier of the replica # of the printed line.
- metal: Each plate will have up to six metal column headers, where the possible metals include: ['Ce', 'Co', 'Cu', 'Fe', 'In', 'Mg', 'Ni', 'Sn', 'Ta', 'Y']. The metal columns sum to 1, indicating the ratios of metals printed.

ii. The data in aggregated_replicas/ consists of optical transmission data, with one row per tenfold aggregated patch on a plate. The column headers are:
- signal_: The measurement of \( \alpha \), the optical transmission spectrum of the printed patch, at a given wavelength. \# ranges from 0 to 8, inclusive, indicating transmission spectra at the following wavelengths: 375, 395, 455, 530, 590, 617, 660, 735, & 850 nm.
- plate: The integer plate identifier.
- line: An integer identifier of the composition gradient that was printed.
- line_experiment_id: An integer identifier of the composition sample along the composition gradient.
- metal: Each plate will have up to six metal column headers, where the possible metals include: ['Ce', 'Co', 'Cu', 'Fe', 'In', 'Mg', 'Ni', 'Sn', 'Ta', 'Y']. The metal columns sum to 1, indicating the ratios of metals printed.

iii. The data in mixture/ represents the outcome of a probabilistic model that a given composition can be explained by a mixture of at most 3 binary signals. There is one row per composition. The column headers are:
- log_prob: The log of the probability that this composition is explainable by at most 3 binary signals.


- metal: Each plate will have up to six metal column headers, where the possible metals include: ['Ce', 'Co', 'Cu', 'Fe', 'In', 'Mg', 'Ni', 'Sn', 'Ta', 'Y']. The metal columns sum to 1, indicating the ratios of metals in the composition.

iv. The data in phase_fits/ represents the outcome of a phase fitting model. There is one row per phase diagram. This data is meant to be read using the example colab (https://colab.research.google.com/drive/1Z5MXnBuC9Y5Nj6rYSulSeCejKyRQGmuo). The column headers are:
  - residual: Float, the residual of the phase fit.
  - signal_type: This is either 'signal' or 'sigma', indicating the type of the phase fit (see paper).
  - discretization: The integer number of intervals we discretized the phase space into.
  - n_points: The number of internal points in the phase diagram. This is an integer between 1 and 5, inclusive.
  - metal_0, metal_1, metal_2: Three strings identifying the constituent metals of the phase diagram.
  - point_#_pos_0, point_#_pos_y: The coordinates of a phase point. # ranges between 0 and 7, inclusive. point_#_pos_0 gives the float amount of metal_0, and point_#_pos_1 gives the float amount of metal_1. The float amount of metal_2 can be inferred via 1 - (point_#_pos_0 + point_#_pos_1).
  - point_#_fitted_channel_X: The fitted optical absorption spectra of point_. # is an integer between 0 and 7, inclusive. X is an integer between 0 and 8, inclusive, indicating the wavelength of the light absorbed.

The files are publicly available for access via:
- the gsutil CLI tool at https://cloud.google.com/storage/docs/gsutil
- After installing the Google Cloud SDK, the ca. 1.3 GB dataset can be downloaded with the command "gutil cp -r gs://gresearch/metal-oxide-spectroscopy <local path>" where <local path> is the target destination for creating the metal-oxide-spectroscopy folder on your local drive
- the tf.io.gfile APIs at https://www.tensorflow.org/api_docs/python/tf/io/gfile/GFile, or
- the HTTP API. Specific filepaths available via http include:
  - the README at: http://storage.googleapis.com/gresearch/metal-oxide-spectroscopy/README.txt
  - the metadata file at: http://storage.googleapis.com/gresearch/metal-oxide-spectroscopy/metadata.csv, which lists all the plates available for download, and
  - the plate data for each of the four data types listed above can be accessed via a similar link based on the folder structure of gs://gresearch/metal-oxide-spectroscopy
Fig. S1. a) Illustration of a phase diagram with a single 3-cation phase (red circle) with corresponding phase field boundaries shown as red lines. The blue lines indicate composition lines drawn from each end member composition to the midpoint of the opposing boundary of the triangle composition graph. The composition lines intersect the phase boundaries at the indicated points. b) The same concept is shown in 4-component composition space where the phase boundaries are 2-D surfaces and the blue composition lines extend from each end-member to the middle of the opposing face of the tetrahedron composition graph. c) The design of composition lines for comprehensive mapping of optical properties in ternary composition space.
Fig. S2. a) 10 unique composition paths starting from element A and continuously traversing the composition graph with different combinations of 4 composition line segments from Figure S1c. Each of the 4 composition line segments is shown as a color gradient arrow starting with red and then proceeding to orange, yellow, green and blue. b) The projection of these 4 composition line segments (gradient-colored arrows) onto a single printed composition line. c) The full design of the printed composition library containing 84 printed lines, each containing 4 composition line segments, which collectively include 10 duplicates of each of the 48 unique composition line segments.
Fig. S3. A demonstration of the concept introduced in Figure S1. The optical signal plotted along the composition lines (left) is shown alongside the corresponding fitted phase diagram (middle, phase field boundaries shown as blue lines). The enlarged composition line in a) is shown in gray in the phase diagram, with the 4 intersections with phase boundaries shown as number stars c) The full 9-channel optical signals along the composition line are shown with the positions of the phase boundary crossings. A change in slope in the optical signals is observed at each phase boundary crossing, with star #2 having the most subtle change due to the composition line and phase boundary line being nearly parallel.
Fig. S4. Representative examples of the plate-level quality control from workflow step 8 (see Figure 1). The points correspond to the discretization of composition gradient lines, and the lines of points correspond to the print design of Fig. S2c. The points are colored according to the universal color scale shown between c) and d), which is the absorption signal difference between that measured at each point and the median over the 10 printed duplicates on the plate, all for the central channel of the 9-channel measurement. The 3 examples shown here represent a) no discernable synthesis artifact, b) a crack in the glass substrate that is not impactful for the aggregate signal but does impact the noise model, and c) a gradual spatial artifact that is not impactful for the aggregate signal and leads to an increased $\sigma$. The plate in b) was flagged as invalid while the plate in c) proceeded through the analysis workflow, a choice validated by the smooth composition trends in aggregate optical signals shown in d).
Fig. S5. XRD characterization in the Ta-Co-Fe system. The Ta-Co compositions and Fe-Ta compositions exhibit the noted 2-cation oxide phases with signals notably different than those of the end-member oxides. The identified phases $\text{Ta}_2\text{CoO}_6$ and $\text{FeTaO}_4$ have overlapped signals, circumventing definitive identification of the phase or phase mixture in the pair of measured 3-cation compositions.
Fig. S6. XRD characterization of the Cu-In-Fe oxide system annealed at 550°C. The results include a) optical data and b) model phase diagram from the high throughput workflow, from which the predicted phases for select compositions are shown in c). This table also includes the results for identified phases where the phase labels are defined in d). The XRD data supporting these results are provided for e) Cu-Fe compositions, f) Fe-In compositions, and g) Cu-Fe-In compositions.
Fig. S7. XRD characterization in the Sn-Co-Ta oxide system where all 5 of the 3-cation compositions have a similar XRD pattern with insufficient intensity to distinguish between the similar patterns of TaO₂ and Ta₂CoO₆. Multiple duplicates of the 3-cation compositions were measured to demonstrate reproducibility.
Fig. S8. Optical characterization of sputter-deposited Sn-Co-Ta oxides. The spectral a) transmission and b) reflection spectra were used to model c) the absorption coefficient. The legend shows that each spectrum is colored according to the cation composition, although the composition trends are most evident in the evaluation of the absorption coefficient at 3 representative energies in Fig. 4c. Note that the high transmission samples exhibit interference fringes in a) and b).
Fig. S9. Ellipsometry characterization of 3 compositions from the sputter-deposited Sn-Co-Ta oxide library. a) Absorption coefficient from fitted ellipsometry measurements, where Co-rich samples exhibit appreciable absorption in the upper-visible spectrum and the 7% Co sample exhibits low optical absorption across the visible spectrum. (b-d) Ellipsometry data and model fit for each sample.
Fig. S10. Additional electrochemical and corrosion data from the analysis of sputter-deposited Sn-Co-Ta oxides. OER experiments in both pH 3 (left, middle) and pH 0 (right) show similar composition trends in the current density observed at 1.72 V vs RHE after 200 s of operation (top) as well as in the relative activity at 200 s with respect to 20 s of operation (bottom). The left-most image summarizes the activity-stability trade-off in pH 3 electrolyte, where the red diamond corresponds to a CoO₅ thin film catalyst and the remainder of the points are compositions from a Co-Ta-Sn oxide library deposited on a conductive Pt support. The gray shaded region corresponds to current densities near the noise floor of the measurement. The dashed line is a visual guide of the approximate Pareto front for co-optimization of maximum activity and minimum corrosion.
Fig. S11. Illustration of the discretizations of 3-cation composition space for computational modelling. a) The grid of compositions with \( \frac{1}{6} \) intervals for the phase diagram model, where the 3 vertices are always considered as phase points and additional phase points are added at one of the composition points shown here. b) The grid of compositions with \( \frac{1}{10} \) intervals for the emergent property model. Perimeter compositions are shown in black and interior compositions are shown in grey.