

Benchtop Autonomous Electrochemical Characterization System for Combinatorial Thin-Film Solid Oxide Electrodes

Jake D. Huang^{1,2,*}, Meagan C. Papac², Davi M. Fébba², Ryan P. O'Hayre¹,
and Andriy Zakutayev²

¹*Metallurgical and Materials Engineering, Colorado School of Mines, Golden, Colorado, USA*

²*Materials Science Center, National Laboratory of the Rockies, Golden, Colorado, USA*

 (Received 29 May 2025; revised 27 January 2026; accepted 9 February 2026; published 13 March 2026)

The design of materials for electrochemical energy conversion is complicated by a vast search space of candidate materials and multifaceted property requirements: multicarrier conductivity, stability, and catalytic activity are all necessary but rarely intersect. Although self-driving laboratories are rapidly rising to address such material optimization problems, the required infrastructure for integrated, large-scale robotic facilities can be cost-prohibitive. Here we develop and evaluate a closed-loop measurement system for efficient screening of proton-conducting oxide electrodes for ceramic fuel cells and electrolyzers, building on top of an existing benchtop instrument and integrating techniques for rapid impedance measurement and automated analysis. This system exemplifies a “minimum viable” self-driving implementation that can deliver substantial benefits with relatively simple infrastructure. Combinatorial thin-film micro-electrode libraries are characterized with a recently developed joint time-domain and frequency-domain impedance measurement technique, which provides an order-of-magnitude acceleration relative to conventional impedance spectroscopy. The distribution of relaxation times is extracted from impedance data and analyzed without human intervention. These results feed an active learning and Bayesian optimization process that learns to predict electrochemical impedance as a function of material composition, measurement temperature, oxygen partial pressure, and electrical bias, which further reduces the screening time by tenfold with optimized experimental sequences. We apply this system to $\text{Ba}(\text{Co}, \text{Fe}, \text{Zr}, \text{Y})\text{O}_{3-\delta}$ combinatorial libraries and evaluate its effectiveness for learning material property trends and optimizing expensive-to-evaluate properties such as activation energy. This offers insights into key methodological aspects of practical autonomous experimentation, including surrogate model validation, cost-aware acquisition functions, and high-throughput data interpretation. Our results demonstrate the efficacy of the system for rapidly gathering information, but also highlight real-world experimental challenges of thin-film degradation and numerical instability in surrogate models.

DOI: [10.1103/nkhy-y2nz](https://doi.org/10.1103/nkhy-y2nz)

I. INTRODUCTION

Material design and discovery is entering a new era in which sophisticated computational and experimental tools can substantially enhance the abilities of human researchers. High-throughput computational screening can quickly provide a set of candidate materials that are likely to be stable and/or provide desired properties [1–4].

Meanwhile, high-throughput experimental systems can quickly screen materials via rapid synthesis and characterization [5–7]. By the pairing of recent developments in artificial intelligence with computational tools and automated experimental systems, the vision of self-driving laboratories is now much nearer to reality [8–10]. Successful demonstrations already exist, e.g., for combustion synthesis of palladium films [11], optimization of optoelectronic properties in organic thin films [12], and inorganic powder synthesis [13]. Self-driving laboratories are rapidly growing in diversity and capability, as evidenced by the emergence of ambitious frameworks leveraging fully robotic laboratories [14] and sophisticated distributed infrastructures [15,16]. However, given that not all institutes can build large-scale, capital-intensive autonomous laboratories, it is also worthwhile to deploy small, modular autonomous systems to

*Contact author: jdhuang@mines.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by National Renewable Energy Laboratory (NREL) Library, part of a national laboratory of the U.S. Department of Energy.

harness expanding artificial intelligence tools in ordinary facilities.

Within the ecosystem of autonomous materials research, materials for solid-state electrochemical energy storage, such as solid-state batteries or solid oxide fuel cells, have received relatively little focus. While conventional solid-state inorganic synthesis is challenging to automate [8,13,14,17], chemical and physical vapor deposition provide a relatively facile route to programmable synthesis of thin films. Combinatorial deposition can produce compositionally graded films that span broad chemical spaces [18,19]. Several recent studies demonstrate autonomous or semiautonomous deposition of inorganic thin films with targeted compositions or properties [20–23]. These automated synthesis routines provide, in principle, a foundation for closed-loop electrochemical materials screening. Nonetheless, understanding the intricate processes in these materials, such as ion and electron transport, interfacial charge transfer, and surface reactions, is challenging. Electrochemical impedance spectroscopy (EIS) is a valuable analytical technique, but is limited by the need for time-consuming measurements under various conditions and ambiguous, manual data analysis [24,25]. Although simulations consistently indicate that techniques such as Bayesian optimization can accelerate materials design [26], real-world autonomous experimentation brings additional challenges such as sample degradation and anomalous measurements. Thus, characterization and data interpretation remain a challenge in autonomous electrochemical materials screening [6].

Recently, we developed a benchtop system for high-throughput electrochemical measurements of thin-film microelectrode libraries fabricated via combinatorial pulsed laser deposition (PLD) [19,27]. The perovskite oxide $\text{BaCo}_x\text{Fe}_y\text{Zr}_z\text{Y}_{1-x-y-z}\text{O}_{3-\delta}$ (BCFZY), a promising cathode material for proton-conducting ceramic fuel cells [28,29], was synthesized via PLD with varying compositions. Combinatorial libraries of BCFZY electrodes then underwent EIS characterization by a robotic electrochemical probe at different temperatures and atmospheres. A multidimensional equivalent-circuit model was developed to analyze the resulting data, revealing trends in electrode performance with composition. However, exhaustive EIS measurements over all electrodes, temperatures, and atmospheres required several days of measurement time per library, while manual data analysis required substantial effort and additional time investment. Our recent work in solid oxide fuel cells at the button-cell scale offers tools to mitigate these limitations: a current-interrupt technique enables accelerated impedance measurement [30], while robust algorithms enable self-tuning distribution of relaxation times (DRT) transformation of impedance data [31,32].

Here we implement a “minimum viable” autonomous workflow for efficient characterization and screening of

combinatorial microelectrode libraries, integrating the existing instrument with rapid EIS measurement and automatic DRT transformation techniques. We develop a self-driving PYTHON engine to efficiently determine property trends and autonomously identify property optima, replacing the previous automated grid measurement. The rapid EIS technique accelerates impedance measurement 20-fold, while the DRT provides a basis for facile interpretation without a human in the loop. Autonomous data collection and analysis enables active learning experiments to efficiently identify trends in electrode performance, using the DRT as a vector property that provides insight into the processes that contribute to electrode losses. Examination of the same BCFZY libraries under an expanded range of measurement conditions provides new insight into electrode processes and nuances of the microelectrode stack configuration. We examine the efficacy of variance-based and cost-aware learning algorithms, and apply principal component analysis (PCA) and partial dependence analysis to extract physicochemical insight from the active learning results. In addition, a Bayesian optimization algorithm for identifying the electrode composition with the minimum activation energy is implemented. Although the connection between thin-film screening and real-world utility should be studied further, our results highlight Co-rich BCFZY compositions as promising electrode materials, in good agreement with demonstrated device performance [28,33]. We develop a variety of metrics to evaluate both active learning and Bayesian optimization experiments, which highlight both strengths and weaknesses of the closed-loop process. We examine our results critically, illustrating how autonomous processes can accelerate electrochemical data acquisition and analysis, while also discussing the challenge of sample degradation during prolonged measurements and opportunities for further development of autonomous approaches.

II. METHODS

A. Synthesis and instrumentation

We give a brief overview of the experimental sequence of the study here and in Fig. 1; detailed descriptions are given in Sec. S1 in Supplemental Material [34]. Three thin-film libraries, each comprising a 12×12 grid of BCFZY microelectrodes, were synthesized with varying composition at the B site of the ABO_3 structure via combinatorial PLD as described previously [19]. The crystal structure and chemical composition of the libraries were determined by spatially resolved x-ray diffraction and x-ray fluorescence (Sec. S.2 in Supplemental Material [34]). The composition gradient for library 1 is illustrated in Fig. 1(a), while quantitative composition maps for all three libraries are shown in Fig. S1 in Supplemental Material [34].

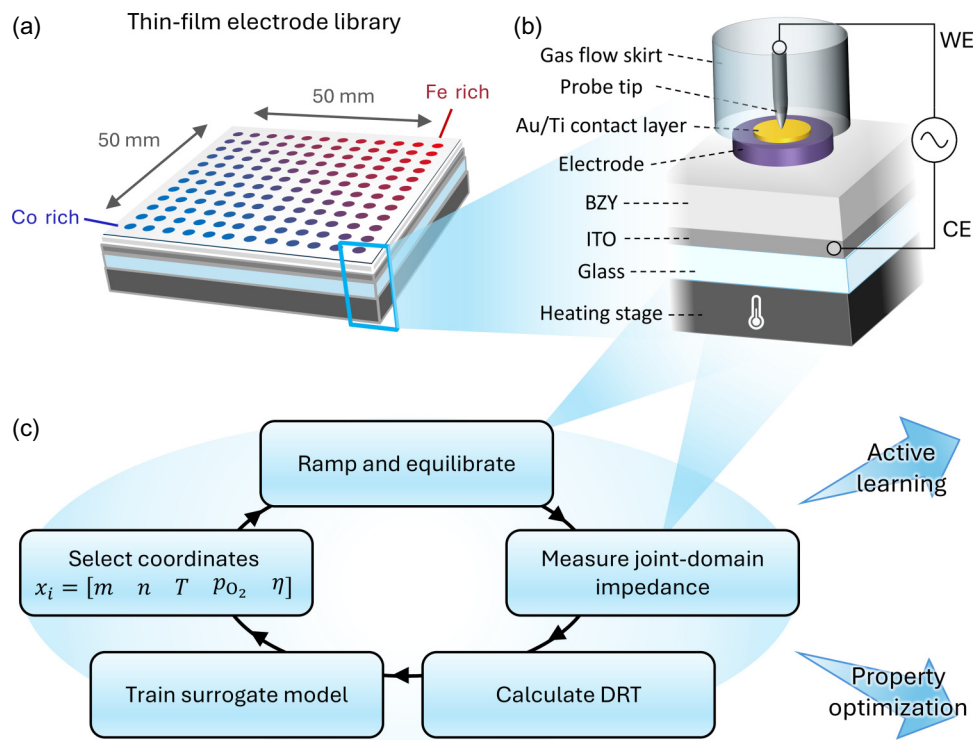


FIG. 1. Overview of autonomous measurements of microelectrode libraries. (a) Each library consists of a 12×12 grid of microelectrodes with differing composition. Library 1 is displayed here, with color illustrating the variation in B-site composition. (b) A robotic probe measures out-of-plane impedance at each microelectrode, with BCFZY acting as the working electrode (WE), BZY acting as an electrolyte layer, and ITO serving as the counter electrode (CE). A thin Ti/Au layer on top of BCFZY ensures electrical contact with the probe. A glass “skirt” contains controlled gas flow at the probe tip to maintain the local atmosphere of the microelectrode under measurement. (c) An autonomous measurement and analysis cycle is used to investigate each library, allowing closed-loop active learning and property optimization. See also Fig. S2 in Supplemental Material [34] for a photograph and schematic of the instrument.

Each microelectrode is part of a layered stack of materials, as shown in Fig. 1(b), which serves as an electrochemical half cell. A robotic electrochemical test station, which uses an x - y -translating heating stage and a z -translating probe arm to access individual microelectrodes, was used to perform out-of-plane measurements at controlled temperature and atmosphere (Sec. S1.3 in Supplemental Material [34]) [19]. For these measurements, indium tin oxide (ITO) acts as the counter electrode, $\text{BaZr}_{0.8}\text{Y}_{0.2}\text{O}_{3-\delta}$ (BZY) as a proton-conducting electrolyte, and BCFZY as the working electrode, while the Ti/Au layer ensures electrical contact with the probe tip, as shown in Fig. 1(b).

B. Autonomous electrochemical measurements

The electrochemical measurement system is controlled by an autonomous experimentation engine implemented in PYTHON, described in detail in the following sections and in the Appendix. The engine iteratively selects, executes, and analyzes impedance measurements to develop a surrogate model of the library under measurement

[Fig. 1(c)]. The PYTHON engine controls the entire flow of the experiment via bidirectional communication with all hardware, including x - y - and z -translating stages, a heating stage, mass flow controllers, and a potentiostat (image and schematic shown in Fig. S2 in Supplemental Material [34]). At each iteration, the engine selects a set of coordinates x_i in the experimental search space, which is described by the electrode position (row and column), film temperature T_{film} , oxygen partial pressure p_{O_2} , and applied dc voltage bias η (Sec. II B 1). The measurement conditions are adjusted to the new coordinates, and the sample is allowed to equilibrate (Appendix A 1), after which an accelerated impedance measurement is performed at these coordinates, and the measurement is transformed to the property of interest (Sec. II B 2). The newly measured property value (in combination with all preceding measurements) is used to train a Gaussian process (GP) surrogate model, which learns the property as a function of the search space coordinates (Appendix A 2). The surrogate model then drives autonomous experimentation in one of two different modes: active learning or Bayesian optimization (Sec. II B 3).

1. Experimental search space

Experimentation involves the exploration of a search space. Material discovery efforts, generally speaking, are concerned with the exploration of the space of candidate materials for a particular application. High-throughput combinatorial synthesis addresses this space by rapidly generating candidate materials. However, this work focuses on the characterization of individual combinatorial libraries, in which the space of candidate materials is constrained by the available (synthesized) electrode compositions. We seek to learn not only how properties vary with composition within the library, but also how they vary with measurement conditions, which can provide insight into the underlying mechanisms. Thus, for the characterization of a single electrode library, we consider the search parameters to be the film composition (determined by electrode row index m and column index n , set by the x - y stage position), the heating stage temperature T_{stage} (resulting in surface temperature T_{film}), oxygen partial pressure p_{O_2} (determined by the mixing ratio of air and nitrogen, set by two mass flow controllers), and dc electrical bias η (applied by the potentiostat). The range of values explored for each dimension is summarized in Table S1 in Supplemental Material [34]. Any point in the search space can then be expressed in terms of coordinates x :

$$x = [m \quad n \quad T_{\text{stage}} \quad p_{\text{O}_2} \quad \eta]. \quad (1)$$

If one wished to perform exhaustive grid measurements over this search space with only five grid points per condition dimension, $5^3 \times 144 = 18\,000$ measurements would be required per library, which is clearly impractical.

2. Target properties and measurement approach

Electrode materials must provide a suitable combination of charge transport and catalytic activity to facilitate the desired electrochemical reaction. The polarization resistance R_p of the electrode can be viewed as a proxy for the electrode's effectiveness in this function: a lower R_p value indicates a more facile reaction at the electrode. Although the operational performance of the electrode depends on cumulative losses (the integral of R_p over current density), R_p is a reasonable metric of electrode performance. Frequency-dependent impedance provides greater insight into this metric: the relaxation of distinct processes at different characteristic frequencies can identify major contributions to R_p . The DRT enables model-free, intuitive analysis by transforming the impedance spectrum to a distribution over time constants. Peaks in the DRT can generally be interpreted to represent different physicochemical processes [35,36]; their time constants and resistances are indicated by peak position and integrated area, respectively. Meanwhile, R_p is given by the integral of the DRT

across all timescales. Thus, we use the DRT as an informative representation of impedance for active learning and optimization experiments, as it both captures the overall metric of performance (R_p) and provides information about the underlying processes.

EIS measurements were used to probe the electrodes under various conditions. To reduce the time cost associated with conventional impedance measurements (approximately 8 min per sample for the frequency range from 100 kHz to 50 mHz), joint-domain "hybrid" impedance measurements [30] were implemented to accelerate data collection 20-fold, enabling measurement over the same frequency range in just 20 s. A robust self-tuning inversion algorithm [32] was used to transform the joint-domain data to the DRT, allowing direct R_p calculation subsequently.

3. Autonomous experimentation modes

Two different modes of autonomous operation are considered: active learning and Bayesian optimization. Both modes rely on a machine-learned surrogate model (Appendix A 2), which is trained to predict the target property as a function of search space coordinates. At each iteration, the surrogate model is evaluated with acquisition functions that drive the selection of a new set of coordinates to measure.

In active learning experiments, the engine is directed to learn the DRT as accurately as possible using a minimum number of measurements as a function of all five search space dimensions, providing basic physicochemical insight. In an active learning experiment, we seek to maximize the predictive accuracy of the surrogate model by using acquisition functions that prioritize uncertainty reduction. We compare two acquisition functions: point-wise variance reduction (PVR), which evaluates surrogate model uncertainty only at the test coordinates, and integrated variance reduction (IVR), which considers uncertainty throughout the entire search space (see Appendix A 3 for details).

In Bayesian optimization experiments, the engine is directed to identify the optimum of a target property within the search space in a minimum number of measurements. In contrast to active learning, Bayesian optimization may not accurately determine properties throughout the entire search space, but rather seeks to quickly identify the region of the property optimum. Instead of simply targeting R_p , we seek to minimize the activation energy E_a , which can be estimated only via temperature-dependent R_p measurements. Fixing $p_{\text{O}_2} = 0.21$ atm and $\eta = 0$ V, we seek the composition (electrode row and column) with lowest E_a under these conditions, which necessitates also exploring a third experimental dimension, T_{film} , to determine E_a as a function of composition. We implement a gradient-predicting GP (Appendix A 2) and a modified probability of improvement (PI) acquisition function

(Appendix A 3) to directly optimize E_a from R_p measurements alone, reducing the number of required measurements and accounting for uncertainty in E_a estimation (see Appendixes A 2 and A 3).

For both active learning and Bayesian optimization, cost-aware variants of the acquisition functions were developed to further increase efficiency by accounting for the time required to change measurement conditions (Appendix A 4). The empirical cost model indicates that the relatively long time required to change the film temperature dominates costs (Fig. S3 in Supplemental Material [34]). To handle experimental anomalies and errors in real time, measurement quality was validated at each iteration, and corrupted or invalid data were flagged and removed by a Bayesian outlier detection procedure (Appendix A 5).

Ideally, all three thin-film libraries would undergo the same set of experiments to generate a comprehensive dataset. However, several experimental challenges prevented this. Library 3 was affected by prolonged high-temperature exposure during initial exploratory measurements and underwent a substantial and systematic change in impedance, likely due to degradation of one or more thin-film layers or interfaces. Although library 3 still produced suitable results for evaluation of active learning approaches, the corresponding data are inconsistent with libraries 1 and 2 (and preceding measurements of library 3), and are thus excluded from the physicochemical analysis presented in Sec. III C. Subsequently, libraries 1 and 2 began to exhibit increasingly frequent electrical contact issues during testing of the Bayesian optimization approach, perhaps due to damage or delamination of the metal contact layer. This became severe enough to preclude meaningful autonomous experiments, such that only library 3 could be evaluated with the finalized Bayesian optimization procedure.

III. RESULTS AND DISCUSSION

A. Measurement validation and reproducibility

To validate the methods used in this work, an independent validation dataset was collected for each electrode library. Each validation set consists of 192 coordinates sampled randomly from the experimental search space. The electrochemical measurement system was programmed to evaluate the randomly selected coordinates using the autonomous ramping and equilibration procedures described in Appendix A 1. Upon equilibration, the system performed an initial hybrid impedance measurement, then a conventional EIS measurement, and finally a second hybrid measurement. The first hybrid impedance measurement (“initial”) is representative of the data that were collected during the subsequent autonomous experiments. The conventional EIS measurement provides a reference spectrum against which the hybrid measurement is validated. The second hybrid measurement (“final”),

which is initiated approximately 8 minutes after the first, is compared with the initial hybrid data to determine whether the sample was fully equilibrated at the initial measurement time. The validation set also provides an independent means of checking the progress of active learning cycles: since the validation set is randomly selected from the experimental search space, it can be used to determine the predictive accuracy of the surrogate model generated by an active learning experiment, as discussed in the following section.

Figure S4 in Supplemental Material [34] displays exemplary data collected during the equilibration period, which illustrates how the stability of fixed-frequency impedance is used to trigger the measurement. Because the electrode and electrolyte layers are thin, the samples equilibrate quickly to changes in conditions and are generally close to equilibrium by the time the controllers stabilize. Small nonmonotonic oscillations during the equilibration period were consistently observed, which likely represent the combination of measurement error and fluctuations in film temperature and gas flow. Approximately 90% of samples equilibrate within 100 s after the controllers stabilize and exhibit impedance shifts of less than 10% during the equilibration period (Fig. S5 in Supplemental Material [34]).

As illustrated in Figs. 2(a)–2(c), the hybrid impedance technique uses EIS to probe high frequencies (100 kHz to 100 Hz) and a chronopotentiometry measurement to probe low frequencies (100 Hz to 50 mHz), and then joins the two measurements via DRT transformation. Despite low-frequency perturbations, the impedance spectrum and DRT extracted from the initial hybrid measurement match conventional EIS data very closely. The second hybrid measurement indicates the same relaxation time constants but a slightly larger resistance than the initial hybrid data [Fig. 2(b)], which is consistent with the magnitude of nonmonotonic oscillation observed during equilibration. Aggregating validation results across all three libraries, we find strong agreement between hybrid measurements (initial and final) and conventional EIS. The coefficient of determination r^2 between conventional EIS and hybrid impedance exceeds 0.99 for more than 99% of measurements [Fig. S6(a) in Supplemental Material [34]]. Meanwhile, the χ^2 goodness-of-fit values obtained by comparing spectra measured via conventional EIS and the hybrid impedance method are consistent with the observed degree of measurement error and run-to-run variability [Fig. S6(b)]. Figure 2(d) compares hybrid initial and conventional impedance spectra corresponding to the 10th, 50th, and 90th percentiles of the χ^2 distribution for the validation datasets for all libraries. While low-frequency perturbation of conventional EIS data is visible, which may be due to film temperature fluctuations or poor contact, hybrid impedance is well-aligned with conventional impedance in all cases. Finally, the R_p values obtained

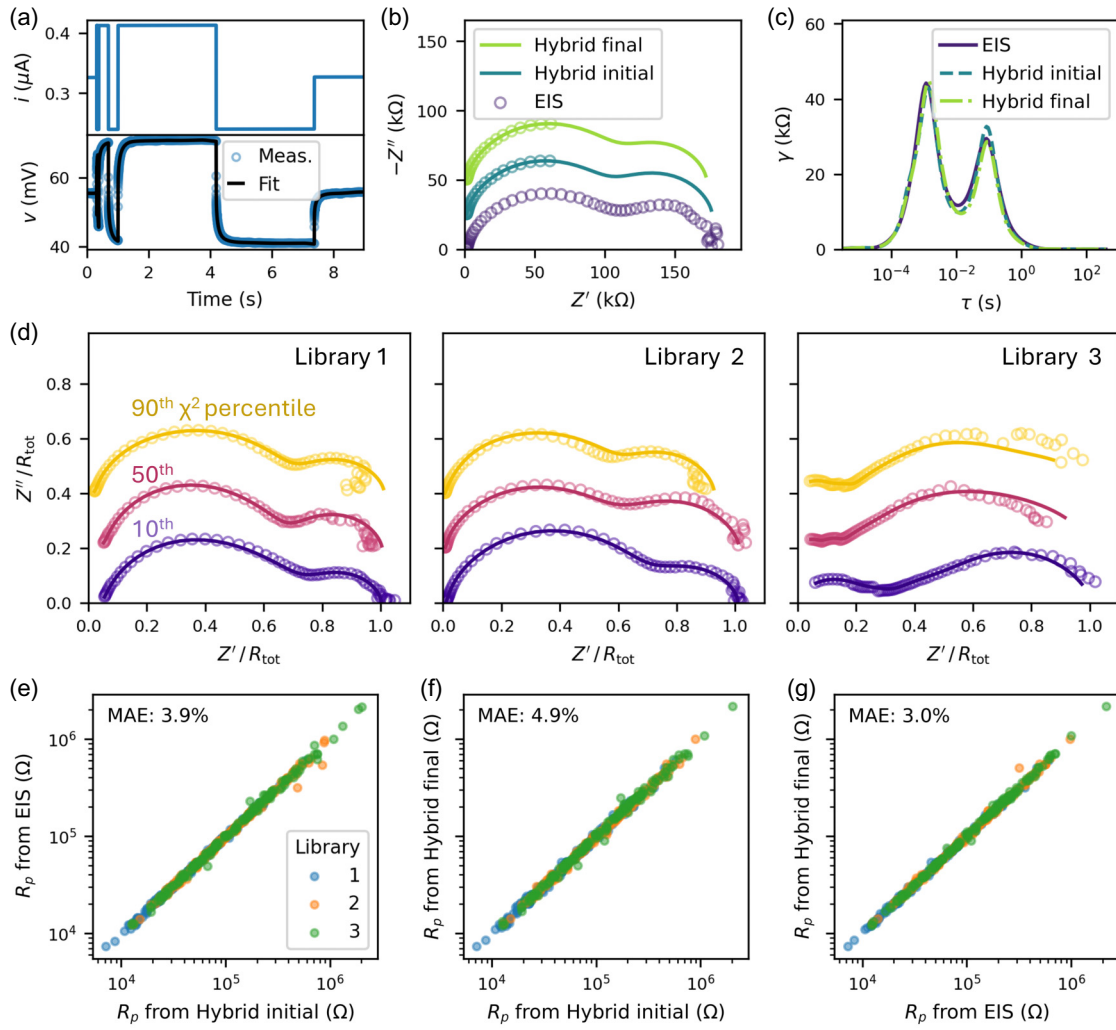


FIG. 2. Measurement validation. (a)–(c) Exemplary hybrid measurement: (a) time-domain component; (b) conventional impedance spectrum compared with impedance spectra extracted from hybrid measurements (initial and final); (c) DRT extracted from conventional and hybrid measurements. In (b), markers for hybrid measurements indicate measured frequency-domain component, while lines indicate impedance extracted from joint-domain data transformation. The measurement displayed corresponds to the 50th χ^2 percentile for library 2 as shown in (d). (d) Comparison of hybrid and conventional impedance measurements from validation datasets. Each panel shows the conventional impedance (markers) and hybrid initial impedance (lines) corresponding to the 10th, 50th, and 90th percentiles of χ^2 values for a single electrode library. In (b),(d), spectra are offset along the y axis for visual clarity. (e)–(g) Comparison of R_p values extracted from different measurements: (e) conventional EIS vs hybrid initial; (f) hybrid final vs hybrid initial; (g) hybrid final vs conventional EIS. MAE, mean absolute error; Meas., measurement.

from all three measurements are in excellent agreement, with no systematic divergence, as shown in Figs. 2(e) and 2(f). Taken together, these data indicate that the electrodes are sufficiently equilibrated at the time of the initial hybrid measurement, and that the hybrid transformation provides an accurate reconstruction of steady-state impedance.

We also investigated the reproducibility of the hybrid measurement and data transformation by examining the initial seed measurements for active learning experiments. A series of eight randomly selected coordinates served as

seed measurements for each active learning run; the same seeds were used for all experiments to provide a consistent initial state. Figures S7–S9 in Supplemental Material [34] compare the results of the nominally identical seed measurements between runs. While there is minor run-to-run variation, the extracted DRT and impedance exhibit high qualitative and quantitative consistency. However, one sample exhibits a severe increase in impedance over time [Fig. S7(c)], suggesting degradation of the film. While most samples remained stable throughout this study, the

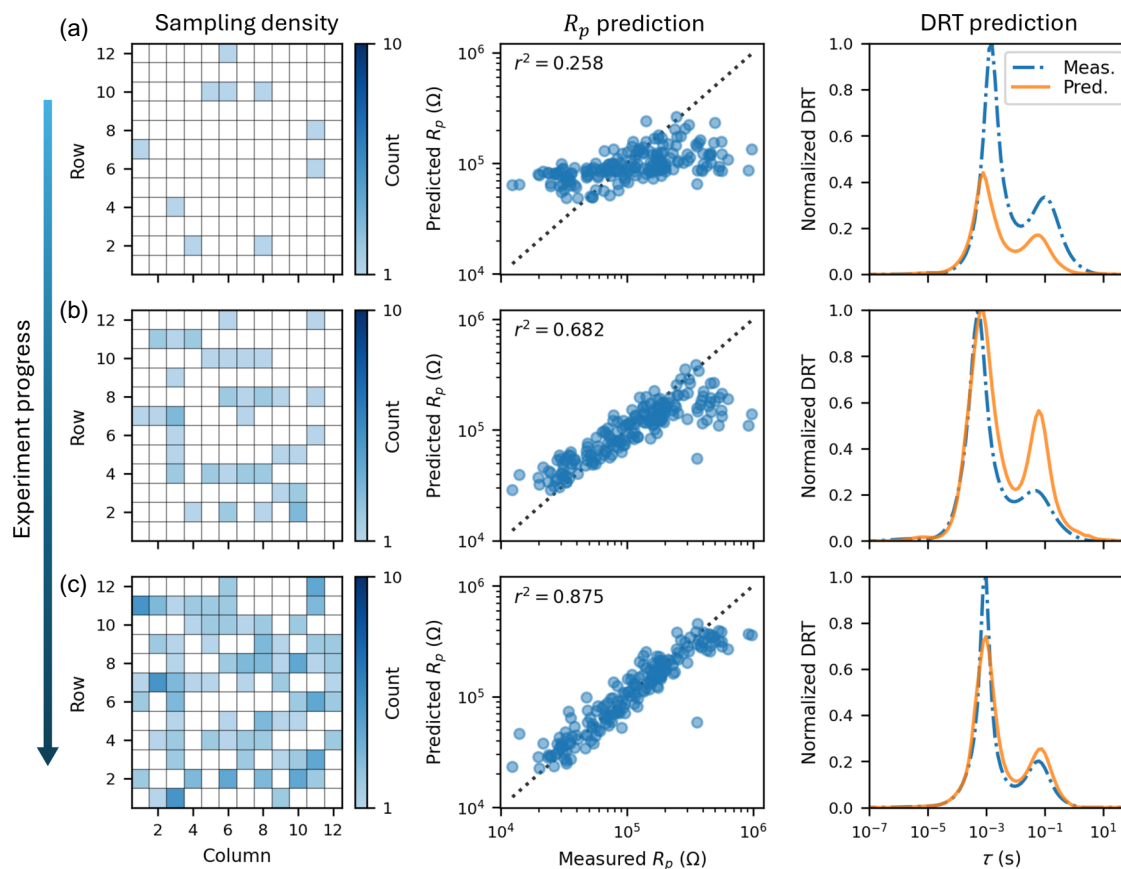


FIG. 3. Exemplary progression of an active learning experiment using the IVR acquisition function. Results are shown after (a) 10, (b) 50, and (c) 150 iterations. Left: Number of measurements taken at each electrode; multiple measurements at the same electrode represent different T_{film} , p_{O_2} , and η coordinates. Center: Surrogate model R_p predictions vs measured values for coordinates in the validation dataset. Right: Surrogate model prediction of the DRT compared with the measured DRT for the validation coordinate corresponding to the median DRT r^2 score at each iteration. Meas., measurement; Pred., prediction.

degradation of some microelectrodes presents challenges (and opportunities), as discussed later.

B. Performance of active learning approaches

1. Definition and evolution of performance metrics

We first evaluate the performance of the active learning approaches tested here. To do so, we first define quantitative performance metrics. Since the objective of these experiments is to determine the landscape of the DRT and R_p as a function of composition and measurement conditions, a primary metric should be the accuracy of the surrogate model in describing the DRT and R_p throughout the search space. Since exhaustive measurements of the entire search space are infeasible, we use the randomly selected validation measurements described in the previous section to evaluate surrogate model accuracy. The surrogate model (trained on active learning experiment data) is used to predict the DRT and R_p for the entire validation set. The R_p prediction accuracy is quantified with the coefficient of determination r^2 . The accuracy of the

DRT prediction for a single coordinate can be quantified with r^2 [32], while the overall DRT prediction accuracy is expressed as the median r^2 value across all validation coordinates. To separate DRT prediction accuracy from R_p prediction accuracy, the DRT is normalized to R_p for r^2 evaluation. Since the number of coordinates contained in the validation set and each active learning experiment is only a small fraction of the search space, the overlap between validation coordinates and active learning (training) coordinates is small (see Sec. S1.5 and Figs. S10 and S11 in Supplemental Material [34]). Thus, the metrics evaluated on the validation set are a reasonable representation of the generalization error of the model (within the search space of the library under consideration).

Figure 3 illustrates the progression of a representative active learning experiment using the IVR acquisition function. After measurement of only ten coordinates, the surrogate model's predictions of both the DRT and R_p are relatively homogeneous and simply reflect the typical values in the initial measurements [Fig. 3(a)]. As the experiment progresses and more coordinates are measured, the

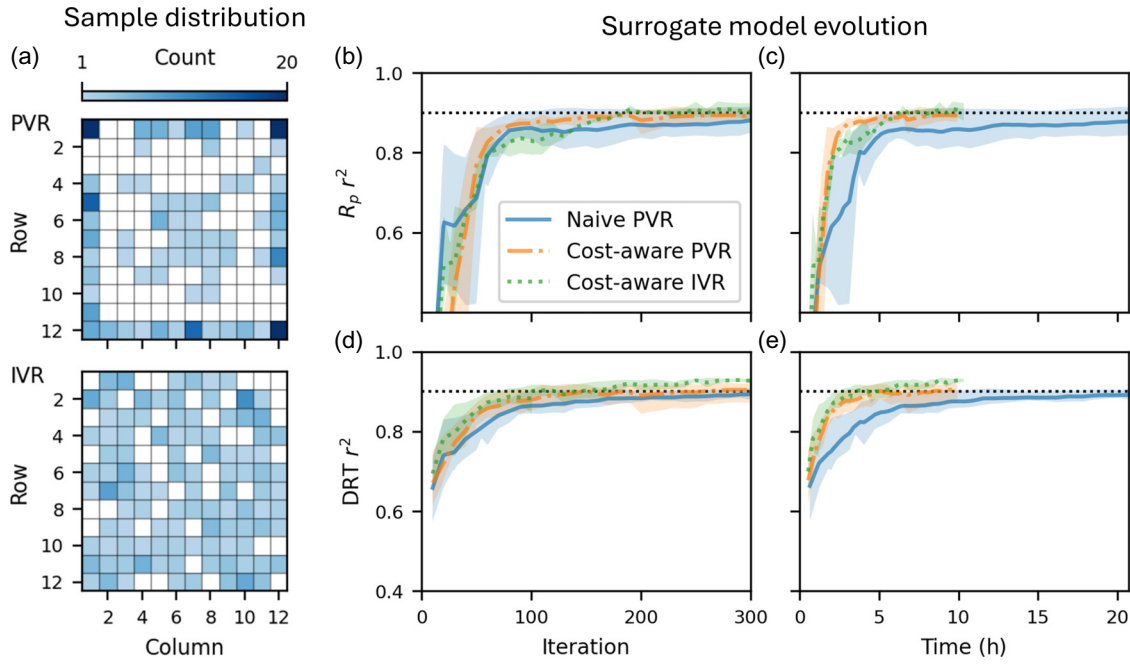


FIG. 4. Comparison of active learning strategies. (a) Number of measurements taken at each electrode for PVR (top) and IVR (bottom) acquisition functions, shown for exemplary experiments after 300 iterations. (b),(c) Surrogate model R_p prediction accuracy vs (b) iteration and (c) time for different acquisition functions. Lines represent the mean of three runs (one run per library), while shaded regions indicate the minimum and maximum values across three runs. (d),(e) Surrogate model DRT prediction accuracy vs (d) iteration and (e) time for the same runs as shown in (b),(c).

surrogate model predicts R_p and the DRT with increasing accuracy, better capturing the trends in R_p and variation in DRT peak sizes and positions [Figs. 3(b) and 3(c)]. The r^2 values can be tracked throughout the experiment, as shown in Fig. S12 in Supplemental Material [34]. We find that across libraries and acquisition strategies, DRT and R_p r^2 values both rise quickly during the first 100 iterations, then tend to plateau at 0.85–0.9, with only marginal improvements throughout the rest of the experiment. The r^2 ceiling at 0.9 can likely be attributed to electrode variations due to deposition artifacts, combined with varying electrical contact quality depending on the exact position of the probe, which lead to unavoidable and unpredictable electrode-to-electrode variability.

We next compare the various active learning strategies tested. While the PVR acquisition function exhibits a strong preference for coordinates at corners and edges of the search space, the IVR function tends to sample the space more uniformly [Fig. 4(a)]. Whereas the PVR is heavily influenced by the inherently higher GP uncertainty at the edges of the search space, the IVR also considers the uncertainty reduction in points surrounding the measured coordinate, which favors interior points. Despite the substantial difference in sampling, both IVR and PVR perform similarly across libraries [Fig. S12(c)]. Interestingly, IVR and PVR also perform similarly to random Sobol sampling, yielding only a slightly faster rise to the R_p r^2

plateau [Fig. S12(d)]. The nearly equal efficacy of all acquisition strategies, despite substantially different sampling, can be explained by relatively uniform property gradients within the search space. Because the DRT and R_p tend to vary smoothly without discontinuities, even random sampling provides sufficient information to enable the surrogate model to reasonably interpolate unseen coordinates. However, this would likely differ in cases where the properties of interest change dramatically between different subspaces (e.g., at phase boundaries). In these cases, acquisition functions that favor discriminatory sampling at the boundaries of such subspaces would be expected to substantially outperform random sampling.

Cost-aware acquisition functions were developed to balance the information gained from a new measurement against the time required to execute the measurement. The introduction of these cost-aware acquisition functions causes the algorithm to change measurement conditions, especially temperature, less frequently to reduce ramping and equilibration times, as shown in Fig. S13 in Supplemental Material [34]. However, cost-aware acquisition functions do not substantially change the evolution of r^2 vs iteration [Figs. 4(b) and 4(d)], indicating that cost-aware coordinate selection does not reduce the incremental information gained via each new measurement. This can be explained by the dominant influence of temperature on cost: as long as the temperature remains fixed, the other

dimensions can be explored relatively freely without incurring large costs, ensuring that information is gained at each iteration despite time expenditure being constrained. However, each cost-aware measurement requires, on average, half as much time as a cost-blind measurement due to the reduction in ramping and equilibration times. As a result, the cost-aware functions accelerate learning by a factor of 2 [Figs. 4(c) and 4(e)] relative to their cost-blind counterparts. The cost-aware functions also provide a compelling speed advantage over random sampling [Fig. S12(h)]. Although random samples could be presorted for efficiency, this would require a predetermined number of samples or batching of random samples, limiting the utility of this approach in the autonomous context.

We also compared discrete and continuous parameterizations of the search space. In the discrete case, each condition parameter (T_{film} , p_{O_2} , and η) was discretized into a grid of five to ten allowed values, whereas in the continuous parameterization, any value in the range could be selected. While the continuous parameterization yields slightly higher r^2 values [Fig. S12(b)], the difference is within the margin of experimental variation, suggesting that both parameterizations are reasonable.

2. Determining experiment completion

Finally, we consider criteria for determining the completion of an autonomous learning experiment. While we used validation sets to evaluate surrogate model accuracy, these validation sets would not be available in the typical context. Thus, a determination of whether the surrogate model is accurate enough must be made on the basis of only the data collected within the experiment. To identify suitable criteria, we compare acquisition function values throughout the course of active learning experiments with accuracy metrics evaluated on the validation sets. As shown in Figs. 5(a)–5(c), there is a strong correlation between DRT and R_p r^2 values (validation set metrics) and the mean and maximum values of the acquisition function at any given iteration of the active learning experiment (available without validation data). While both mean and maximum acquisition function values tend to decrease as prediction accuracy increases, the mean correlates especially well with both r^2 metrics. This suggests that the aggregated acquisition function values can serve as a proxy for surrogate model accuracy.

To demonstrate this, we train a support vector machine classifier to predict whether an active learning experiment can be considered complete. For the sake of this illustration, we define a completion threshold of $r^2 \geq 0.83$ for both DRT and R_p prediction accuracy; i.e., we consider the surrogate model to be sufficiently informative when its prediction r^2 exceeds this value. The threshold value of 0.83 was selected to assign roughly equal numbers of complete and incomplete labels in the training data. The

classifier uses the mean and maximum acquisition function values as inputs to predict whether the surrogate model has reached each of these thresholds. As shown in Fig. 5(d), the probability of experiment completion can then be predicted during the experiment, and the experiment can be terminated once the probability reaches a satisfactory level. The classifier is able to differentiate between well-trained and insufficiently-trained surrogate models with accuracy 0.89 for R_p completion and 0.86 for DRT completion [Figs. 5(e) and 5(f)]. Receiver-operator characteristic curves, shown in Fig. S14 in Supplemental Material [34], further highlight the robustness of the classifiers with area-under-the-curve values greater than 0.9 for both R_p and DRT completion. While the exact correlation between acquisition function values and surrogate model accuracy is likely to depend on the physical system, properties under study, and model architecture, this result suggests that the evolution of acquisition function values during an experiment can be used quite effectively to monitor the learning of the surrogate model and determine when to terminate an experiment.

These results illustrate how the autonomous system can accelerate combinatorial characterization. With cost-aware acquisition and hybrid impedance measurements, the average total time per iteration (ramping, equilibration, and measurement) is 2 min, allowing the surrogate models to achieve a sufficient level of learning in just 5 h (150 iterations). If we assume the same ramping and equilibration times, each iteration would require 10 min with conventional EIS. Applying this to the exhaustive grid of 18 000 coordinates, we find that exhaustive automated EIS would require 3000 h (125 days) per library—600 times longer than 5 h with our approach. At the other extreme, one could greatly sparsify the grid, selecting only two values per condition dimension and one quarter of the electrodes, reducing the number of measurements to $2^3 \times 144/4 = 288$, which would then require only 48 h. However, such a sparse grid would still be 10 times slower, and likely provide coarser information than obtained from the surrogate models here. Thus, it is reasonable to estimate that the autonomous workflow with hybrid EIS and cost-aware acquisition provides at least an order-of-magnitude acceleration relative to automated grid measurement. However, these results may not be transferable to all systems. The radial basis function (RBF) kernel used here (Appendix A 2) is amenable to the smoothly varying DRT in the libraries studied here, but might perform poorly for materials with different phase regions. Physics-informed kernels or other model architectures might be necessary depending on the application.

C. Physicochemical insights from active learning

Next, we examine the surrogate models trained by active learning experiments to extract physicochemical insights

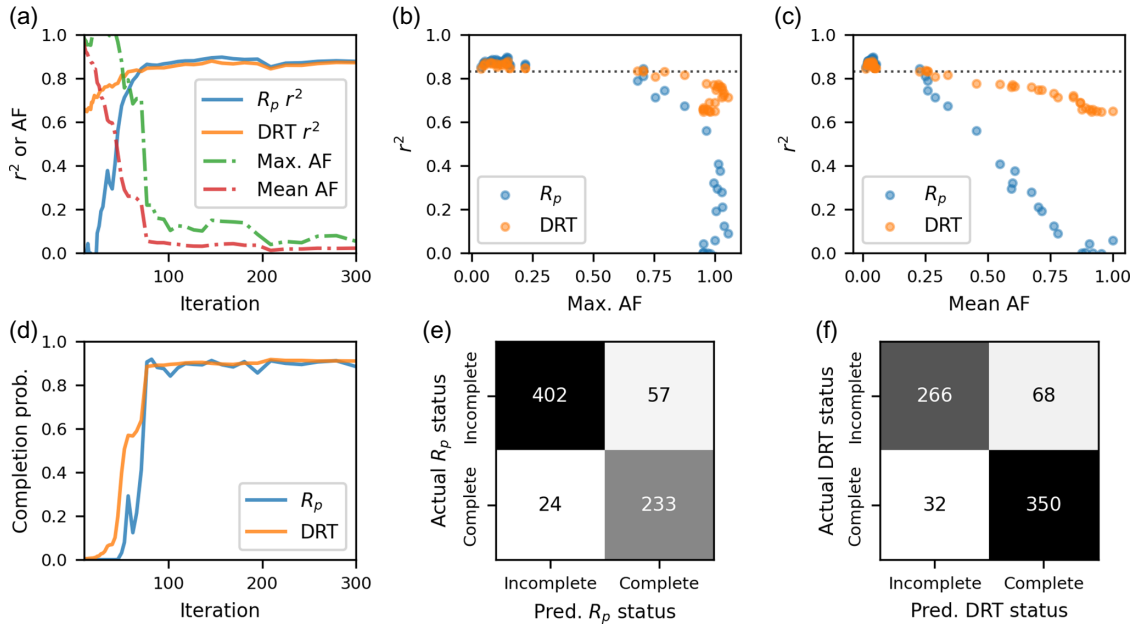


FIG. 5. Active learning stopping criteria. (a) Evolution of acquisition function (AF) values and surrogate model accuracy during an exemplary experiment. Maximum AF (max. AF) and mean AF values at each iteration are normalized to the starting values. (b),(c) Correlation between surrogate model accuracy and normalized AF: (b) maximum AF and (c) mean AF, shown for the exemplary experiment in (a). Horizontal dashed lines indicate the r^2 threshold for experiment completion. (d) Probability of experiment completion for the experiment shown in (a), predicted by a classifier trained on data from all experiments. (e),(f) Confusion matrices for the classifier obtained from cross-validation across all experiments, shown for (e) R_p determination and (f) DRT determination. Pred., predicted; prob., probability.

about the BCFZY material system. For each library, we use the surrogate model trained by the active learning experiment with the highest validation r^2 scores to predict DRT and R_p trends. As noted in Sec. II B 3, library 3 is excluded from this analysis due to anomalous changes in impedance during initial testing.

1. Trends with composition

First, we consider the variation of R_p with composition under fixed conditions ($T_{\text{film}} = 500^\circ\text{C}$, $p_{\text{O}_2} = 21\%$, $\eta = 0\text{ V}$). As shown in Fig. 6(a), R_p values are generally higher in library 2, which occupies the Zr-rich corner of the quaternary B-site composition space. As Zr content decreases and total Co plus Fe content increases, R_p tends to decrease. This is in good agreement with our previous measurements [19]. Meanwhile, the low- R_p region in the center of library 1 (Co plus Fe approximately 80%, Co:Fe ratio 1 or greater, Zr approximately 10% to approximately 20%, Y approximately 0% to approximately 10%) corresponds closely to the well-known, high-performing $\text{BaCo}_{0.4}\text{Fe}_{0.4}\text{Zr}_{0.1}\text{Y}_{0.1}\text{O}_{3-\delta}$ composition [28].

The DRT provides more insight into R_p trends. The mean (composition-averaged) DRT for each library [Fig. 6(b)] indicates two primary peaks that consistently appear at all compositions and conditions: P1, which appears at

$\tau \sim 10^{-4}\text{ s}$ in library 1 and $\tau \sim 10^{-3}\text{ s}$ in library 2, and P2, which appears at $\tau \sim 10^{-1}\text{ s}$ in both libraries. A much smaller peak, likely attributable to proton transport through the BZY layer [19], is typically visible near 10^{-5} s , but tends to be dominated by the shoulder of P1. Since the peak positions and magnitudes vary substantially, as indicated by the ranges in Fig. 6(b), these composition-averaged distributions are only a coarse representation of the typical DRT. To more comprehensively visualize the variation within libraries, Fig. S15 in Supplemental Material [34] shows the DRT as a function of electrode position with respect to the substrate. An additional peak between P1 and P2 is visible for some electrodes, but is typically strongly overlapped with or completely obscured by the primary peaks. This suggests that more than two processes contribute to the electrochemical activity of the electrodes. However, since meaningful separation of additional peaks is generally not possible due to frequency dispersion and peak overlap, we focus on analyzing the two consistently separable peaks, with the caveat that they likely represent multiple process contributions.

To take a deeper look at the influence of composition, we also examine the partial dependence of the DRT on B-site cation fractions. Since the cation fractions are inherently correlated, we use PCA to obtain meaningful partial dependence plots that both obey site balance

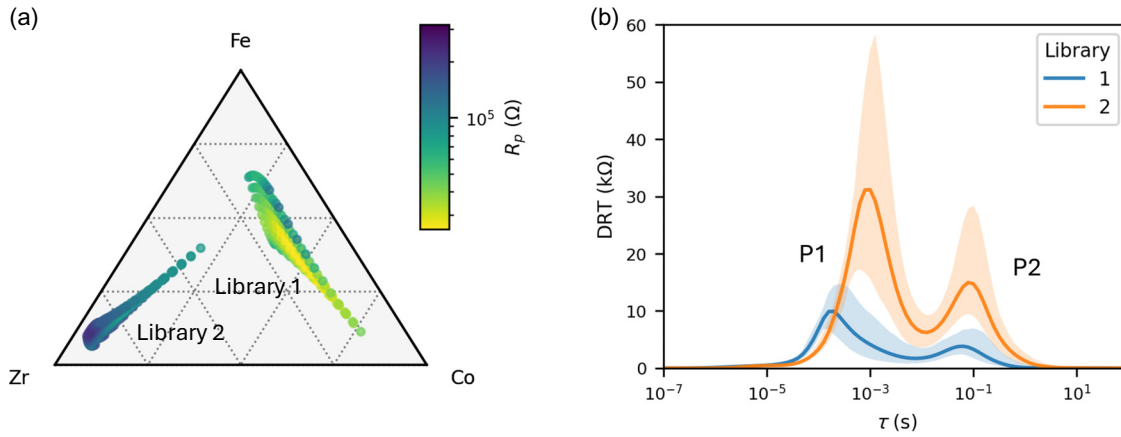


FIG. 6. Trends with composition. (a) Polarization resistance R_p under fixed conditions ($T_{\text{film}} = 500^\circ\text{C}$, $p_{\text{O}_2} = 21\%$, $\eta = 0\text{ V}$) for libraries 1 and 2, displayed as a function of B-site composition. Compositions are visualized in the ternary Co-Fe-Zr plane for simplicity; Y content varies in the range from 0.0% to 6.7% in library 1 and from 5.6% to 24.2% in library 2. Full composition details are given in Fig. S1. (b) Mean DRT (lines) and 10th percentile to 90th percentile range (shaded regions) for each library under the same fixed conditions. Two primary peaks, labeled “P1” and “P2,” are consistently visible in each library.

and reflect the measured compositions. This analysis provides interesting insight into the importance of total Co plus Fe content, the trade-off between Co and Fe, and the importance of Y doping (see Sec. S2.1 and Fig. S16 in Supplemental Material [34]). In addition, PCA results for library 1 indicate that high-Co compositions with Co plus Fe content of approximately 80% tend to exhibit the lowest R_p values [Fig. S16(a)]. This finding is corroborated by a kinetic study of BCFZY compositions with differing Co:Fe ratios, which found that Co-rich $\text{BaCo}_{0.7}\text{Fe}_{0.1}\text{Zr}_{0.1}\text{Y}_{0.1}\text{O}_{3-\delta}$ exhibits high oxygen vacancy diffusivity, fast oxygen surface exchange, and strong performance in both fuel cell and electrolysis operation [33]. While this PCA approach is useful for extracting insights from surrogate models of constrained composition spaces, the interpretability of the data is limited by the outstanding questions regarding the electrode stack discussed in the next section.

2. DRT dependence on measurement conditions

To gain insight into the nature of each DRT peak, we examine the influence of measurement conditions on the DRT. We use partial dependence plots [37], which are commonly used to interpret the influence of individual variables in machine learning models, to visualize the influence of conditions independently. With decreasing temperature, the magnitude of both DRT peaks increases [Figs. 7(a) and 7(d)] as expected for thermally activated processes. The surrogate model identifies a linear relationship between inverse temperature and the logarithm of total resistance (integrated area) of each peak [Figs. S17(a), S17(d), and S17(g) in Supplemental Material [34]]—that

is, the model learns the Arrhenius relationship. Oxygen content has a smaller but noticeable effect: as p_{O_2} decreases, the resistance of P2 increases, while P1 is nearly unchanged [Figs. 7(b) and 7(e)]. This suggests that P2 reflects an oxygen-dependent process or processes, such as a surface reaction step. Finally, the applied dc voltage bias has a surprisingly dramatic monotonic influence. As the voltage moves from negative (oxygen reduction) to positive (oxygen evolution) values, P2 becomes increasingly resistive [Figs. 7(c) and 7(f)]. Under the most oxidizing potentials, a third peak becomes visible in library 1 [Fig. 7(c)] at intermediate timescales ($\tau \sim 10^{-3}\text{ s}$). The trend of P1 with η is less clear: in library 1 (Co and Fe rich), P1 appears to be largely unaffected by η , whereas in library 2 (Zr and Y rich), P1 appears to grow and shift to longer timescales at oxidizing potentials. However, given the appearance of a third peak at $\tau \sim 10^{-3}\text{ s}$ in library 1, this may simply reflect two overlapping processes—one independent of η and the other dependent on η —which both appear near $\tau \sim 10^{-3}\text{ s}$ in library 2.

The typical model of processes in the electrode stacks studied here would include proton transport through BZY (likely assigned to the small, obscured peak near $\tau = 10^{-5}\text{ s}$), mixed hole and proton transport in BCFZY (likely assigned to P1), and oxygen reduction and evolution reactions at the BCFZY surface (likely assigned to P2) [38]. However, the dramatic monotonic influence of η warrants caution, as one would expect the resistance of P2 to decrease with increasing $|\eta|$ due to Butler-Volmer kinetics. To examine polarization effects more closely, we measured the current-voltage relationship for a single electrode, varying η between -200 and 200 mV . We observe diodelike behavior in which current rises rapidly for $\eta < 0$ but asymptotically for $\eta > 0$ [Fig. S18(a) in

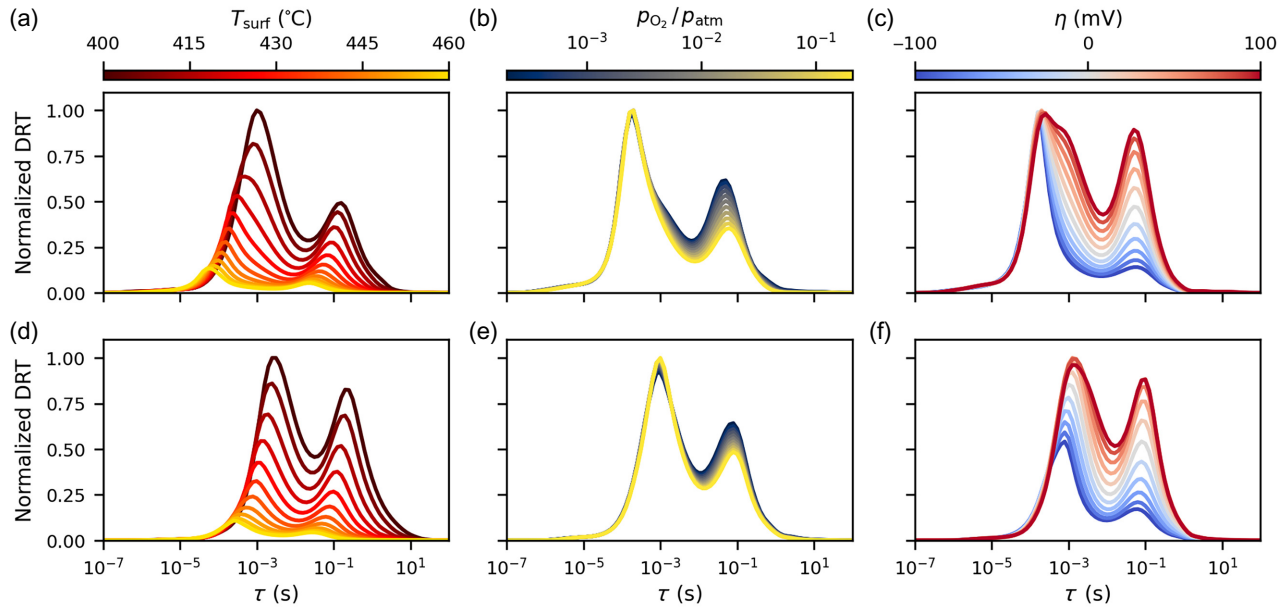


FIG. 7. Partial dependence of the DRT on measurement conditions. Results for libraries 1 and 2 are shown in (a)–(c) and (d)–(f), respectively. Partial dependence is shown for (a),(d) temperature, (b),(e) oxygen partial pressure, and (c),(f) applied dc voltage. In each panel, the DRT is normalized to the maximum calculated value for consistent visual scaling.

Supplemental Material [34]], while differential resistance increases exponentially with η [Fig. S18(b)]. There are many potential influences to consider when interpreting this trend: mass transport limitations, gas adsorption and desorption, potential-dependent defect chemistry, or a possible p - n junction at the BZY/ITO interface. On the basis of the microelectrode stack configuration, observed DRT trends vs bias, corroborating single-electrode bias-dependent impedance (Fig. S19 in Supplemental Material [34]), and current-voltage measurements under differing atmospheres (Fig. S18), we suggest that the observed diodelike behavior is most likely explained by a Schottky barrier between BCFZY and the Ti/Au contact layer, while other mechanisms can be ruled out (Sec. S.2.2 in Supplemental Material [34]). A Schottky barrier might form at the BCFZY/Ti interface because BCFZY is a p -type semiconductor and Ti has a low work function relative to common current collectors (e.g., Au, Pt, and Ag). This barrier would increasingly impede electron transfer with increasing η , hindering reactions at the BCFZY surface and leading to the observed growth of P2 with increasing η . Nonetheless, given the unexpected results and lack of information about the BZY/ITO and BCFZY/Ti interfaces, further investigation is required to conclusively assign the observed DRT peaks and explain this polarization behavior.

Although the electrode configuration studied here requires further study, the analysis approach is flexible and interpretable. The DRT provides a useful semiquantitative tool for understanding electrode behavior. Peak analysis can provide quantitative insights, but is limited by broad frequency dispersion and peak overlap, which is common

in solid oxide devices. One must also keep in mind that the DRT extracted from the surrogate model is unavoidably biased by the model itself, while partial dependence analysis averages over many coordinates and may smooth out fine details. Nonetheless, major trends in the DRT can be extracted with partial dependence analysis supported by PCA, providing a straightforward means to deconvolute the influence of multiple variables on frequency-dependent impedance. We suggest that such an approach can be applied to quickly and autonomously construct an overall picture of electrochemical behavior, at which point human researchers could enter the loop to identify interesting or unusual results for further investigation (such as the unexpected polarization behavior discussed above). The physicochemical intuition of humans is also invaluable for data interpretation as well as potential creative innovations based on the autonomously generated data.

D. Activation energy minimization

Finally, we consider autonomous property optimization. As explained in Sec. II B, to create a nontrivial optimization problem, we seek to identify the electrode composition with the lowest activation energy, fixing $p_{O_2} = 21\%$ and $\eta = 0$ V. E_a optimization experiments were performed with library 3, which was excluded from the preceding physicochemical analysis due to sample degradation. Nonetheless, the trends within library 3 remained self-consistent, such that it still serves as a suitable test system for the optimization strategy, although the composition-property trends are difficult to interpret.

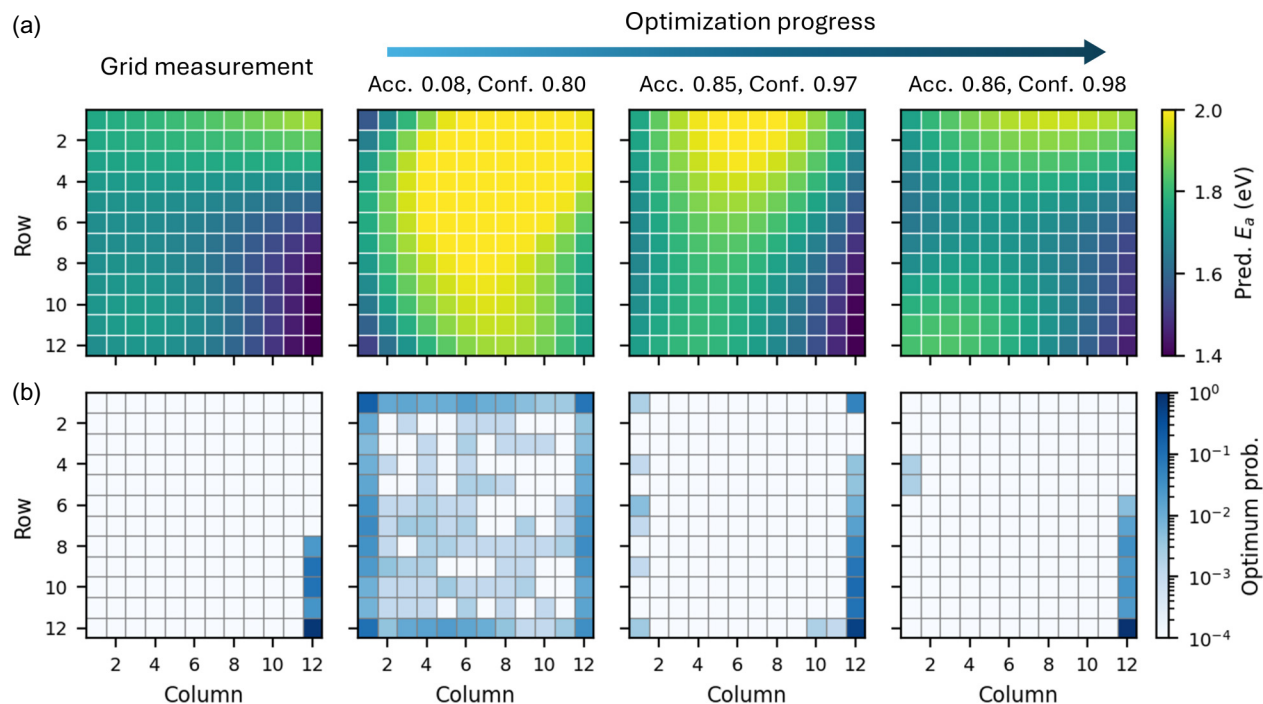


FIG. 8. Evolution of an exemplary activation energy optimization experiment. (a) Predicted activation energy values for all electrodes and (b) predicted probability that each electrode represents the optimum (lowest activation energy). The leftmost column represents values extracted from exhaustive grid measurements, while the remaining columns represent surrogate model predictions after 10, 50, and 100 iterations (left to right). Acc., accuracy score; Conf., confidence score; Pred., predicted; prob., probability.

For validation purposes, exhaustive grid measurements were first performed at every electrode from 450 to 550 °C in 25 °C increments. Activation energies were extracted by fitting the measured R_p values with the composition-kernel GP described in Sec. II B. The resulting activation energy estimates, shown in Fig. 8(a), represent a GP themselves (i.e., E_a and the preexponential factor are treated as composition-dependent GPs). As a result, the E_a estimates for all electrodes comprise a multivariate Gaussian distribution. We sample this distribution to obtain the probability that each electrode represents the optimum in E_a , shown in Fig. 8(b). This indicates that the minimum E_a

most likely appears in the lower-right corner of the library, but could exist at any of the five bottom electrodes in the rightmost column. The uncertainty in which electrode is the true optimum arises not only due to uncertainty in the Arrhenius fit of R_p vs temperature but also because some electrode-to-electrode variability originates from PLD and contact artifacts rather than chemical differences.

Similarly to active learning experiments, the evolution of the surrogate model can be evaluated over the course of an optimization experiment. Following the same approach used for the exhaustive data, E_a values and optimum probabilities are extracted from the surrogate model at each

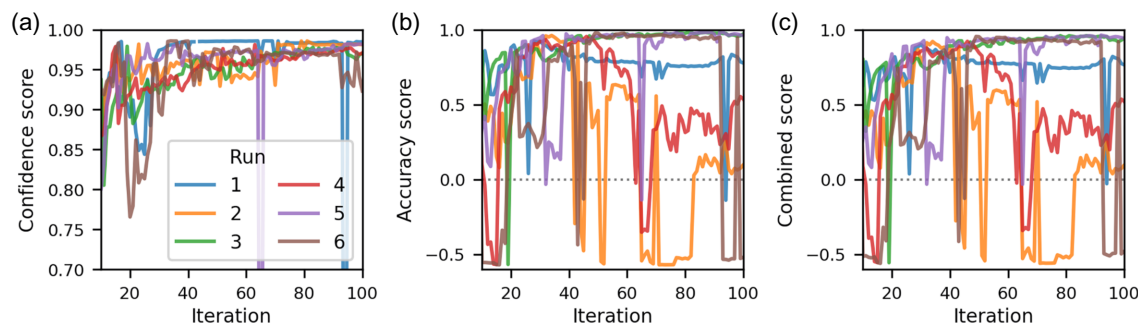


FIG. 9. Performance metric evolution in activation energy optimization experiments: (a) confidence score; (b) accuracy score; (c) combined score (confidence times accuracy). Sudden drops and immediate recovery in score sometimes appear due to numerical instability during training of the surrogate model. In all cases, confidence tends to increase over time, whereas accuracy can decline.

iteration (one iteration represents one R_p measurement), as shown for an exemplary experiment in Fig. 8. At the start of the experiment, the prediction of E_a is inaccurate and the location of the optimum is unclear, as reflected by distributed optimum probabilities. As the experiment progresses, E_a predictions align better with the true values, and the predicted optimum region narrows, ultimately converging with the true optimum region.

As for active learning, we introduce quantitative metrics for the success of E_a optimization experiments, detailed in Sec. S1.6 in Supplemental Material [34]. We quantify the accuracy of the surrogate model on the basis of how well it identifies the optimum region, constructing an accuracy score ranging from 0 to 1 based on the Kullback-Leibler divergence [39]. We quantify the confidence of the model on the basis of the spread of the predicted optimum region (Fig. S20 in Supplemental Material [34]), again defining a confidence score ranging from 0 to 1. Confidence and accuracy scores for six different E_a optimization experiments are shown in Fig. 9. The six runs all use the same optimization approach, but with differing acquisition function configurations, summarized in Table S2 in Supplemental Material [34]. The primary differences include cost-aware vs cost-naïve acquisition functions, alternation between exploitation and exploration acquisition functions, and handling of duplicate coordinate selection. Confidence and accuracy scores are unstable during the first 20–40 iterations, as each new measurement can have a substantial influence, but can stabilize thereafter. For all six experiments there is convergence to reasonably high accuracy scores (0.7 or greater) by iteration 40 [Fig. 9(b)], indicating the effectiveness of the modified PI acquisition function. However, for two of the experiments (runs 2 and 4) there is later divergence to lower accuracy scores. This is most likely due to numerical instability of the GP model when repeated measurements of the same coordinates (or very closely neighboring coordinates) are performed or when invalid measurements (e.g., due to contact loss or degraded electrodes) are not flagged by the outlier detection procedure. Meanwhile, the surrogate model’s confidence rises over time in all cases [Fig. 9(a)] despite the instability in accuracy. To capture the balance of accuracy and confidence, we also create a combined score by multiplying the accuracy and confidence scores, shown in Fig. 9(c). Positive values indicate accurate models, while negative values indicate inaccurate models; in both cases the magnitude of the score increases with confidence. This provides a unified metric for evaluating optimization experiments, as only confident and accurate models can achieve high scores approaching unity.

These results highlight the potential advantage of closed-loop optimization, but also reveal the risk for autonomous agents to develop an incorrect representation of the system under study while reporting a high degree of confidence. We note that the degradation of library 3

causes a large number of corrupted measurements, while E_a extraction from the R_p -trained GP kernel is more sensitive than simple R_p prediction, which creates a challenging setting for optimization. The standard PI acquisition function can also exhibit “greedy” behavior (overprioritization of exploitation) [40], which may contribute to the observed instability. Thus, reliable measurement quality, automated data curation, and stable, noise-tolerant models and acquisition functions are all essential for autonomous optimization. The degree to which each criterion is met determines the reliability of the results. Just as for active learning, periodic human review of optimization results could be valuable for validation and identification of notable trends that may be overlooked by artificial agents.

IV. CONCLUSION

In this work, we develop a system for closed-loop electrochemical characterization of combinatorial thin-film libraries, integrating a benchtop instrument for spatially resolved measurements with techniques for accelerated impedance measurement and robust DRT analysis. This enables efficient active learning of impedance and DRT trends, producing informative surrogate models in a fraction of the time required for conventional grid measurements. We apply the system to BCFZY microelectrodes to understand electrochemistry as a function of B-site composition. Partial dependence analysis of the DRT provides a straightforward means of separating the influence of different search space parameters. Active learning experiments identify anticipated physicochemical trends such as the Arrhenius relationship and the beneficial impact of Co content on performance, but also reveal unexpected polarization behavior. Unfortunately, outstanding questions about the pathway for current flow through the electrode stack hinder deeper physicochemical interpretation. Nonetheless, our results are in good agreement with device studies, highlighting Co-rich BCFZY as a promising electrode material. Bayesian optimization of activation energy using a modified gradient-based acquisition function is effective for quickly identifying the optimum, but highlights the importance of measurement quality and numerical stability.

In comparison with larger integrated systems combining synthesis and characterization, this is a minimal demonstration of self-driving instrumentation. Nonetheless, the benchtop tool provides significant return on a relatively small investment, exemplifying a modular autonomy that is feasible in typical laboratory spaces. The characterization system could be coupled in the future with developing systems for autonomous deposition to create a full synthesis-characterization loop.

Further studies are needed to determine the transferability of thin-film screening to practical device performance. We note that degradation of the thin films caused some

challenges in this study. Since it can be anticipated that samples may degrade over time during long-term high-temperature experiments, it is important to use experimental designs that allow decoupling of degradation from the influence of search space parameters. While a detailed analysis of degradation in this work is stymied by the highly varying treatment histories of different electrodes, it should be possible with careful experimental design to harness thin-film decomposition as a tool for automated or autonomous accelerated degradation studies.

While artificial intelligence is rapidly developing, we still see important roles for human researchers to play in (semi)autonomous experimentation. At this stage, human creativity and physicochemical intuition remains a valuable complement to the automated efficiency and statistical power of autonomous systems. Thus, while self-driving laboratories will no doubt become increasingly commonplace, we encourage researchers to pursue synergy with their artificial counterparts, rather than completely outsourcing scientific thinking. “Human-in-the-loop” workflows could be the best of both worlds, allowing autonomous agents to quickly generate digestible results that can inform and inspire human insight.

ACKNOWLEDGMENTS

This work was authored in part by the National Laboratory of the Rockies for the U.S. Department of Energy under Contract No. DE-AC36-08GO28308. Funding for the measurement instrument and algorithm development and use was provided by the Office of Critical Minerals and Energy Innovation, Advanced Materials and Manufacturing Technologies Program. The impedance method development and synthesis of protonic ceramic materials were supported as part of the Hydrogen in Energy and Information Sciences, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award No. DE-SC0023450.

J.D.H., M.C.P., R.P.O., and A.Z. conceptualized the experiments. J.D.H. developed the methodology and software, performed validation, curated data, produced visualizations, and wrote the original draft. J.D.H. and D.M.F. conducted the experiments. M.C.P., D.M.F., R.P.O., and A.Z. reviewed and edited the text. R.P.O. and A.Z. acquired funding, supervised, and provided resources.

The views and opinions of the authors expressed herein do not necessarily represent those of the U.S. Government.

DATA AVAILABILITY

The data and code that support the findings of this article are openly available [41].

APPENDIX: AUTONOMOUS EXPERIMENTAL OPERATION DETAILS

The electrochemical measurement system is controlled by an autonomous experimentation engine implemented in PYTHON, as described in Sec. II B. The engine uses the PYTHON packages GPyTorch [42] to construct the GP surrogate model, BoTorch [43] to implement acquisition functions, and Ax [44] to manage the experimental search space. Additional details of the implementation are given in the following sections.

1. Ramping and equilibration

At each iteration, a new set of search space coordinates is selected for measurement as described below. The electrode row and column indices, p_{O_2} , and T_{film} coordinates are translated to set points and sent to the corresponding controllers. The controller values are then monitored until they settle at the new set points. To ensure alignment with the electrode, the impedance is measured at 40 Hz to verify electrical contact, and the probe position is automatically adjusted if necessary to achieve electrical contact. The electrode is then allowed to equilibrate to the new conditions before undergoing impedance measurement. Continuous single-frequency impedance measurements at 40 Hz are performed to monitor the equilibration of the electrode. The hybrid impedance measurement is initiated once the rate of change in both real and imaginary impedance declines below 1%/min. The voltage bias η is applied only at the start of the equilibration procedure to avoid prolonged exposure of the film to dc voltages that might promote degradation. The position-dependent film surface temperature is estimated from a model based on infrared images, as described in Sec. S1.4 and shown in Figs. S21–S23 in Supplemental Material [34]. The total gas flow rate was fixed at 100 standard cubic centimeters per minute, which was sufficient to flood the local atmosphere of the electrode while minimizing convective cooling (Fig. S24 in Supplemental Material [34]).

2. Surrogate model

We use a GP surrogate model with an anisotropic RBF kernel, which is a popular choice for active learning and Bayesian optimization [45,46]. This model has several benefits: it is a noise-tolerant, versatile function approximator; the hyperparameters can be automatically selected by maximization of the log-marginal-likelihood to avoid overfitting; and it inherently provides Bayesian uncertainty estimates [47]. Since the oxygen partial pressure is expected to have a power-law or similar influence on electrode processes, p_{O_2} is log-transformed to construct the model input, i.e., $f(p_{O_2}) = \ln(p_{O_2} + 2 \times 10^{-4})$, where the offset of 2×10^{-4} atm prevents infinite values and reflects

the fact that the practical p_{O_2} is nonzero even under pure nitrogen flow.

To model the DRT for active learning experiments, we first discretize the distribution over a grid of time constants ranging from $10^{-7.5}$ to $10^{2.5}$ s with log-uniform spacing of 10 points per decade, resulting in a vector property with 101 elements. We then construct a multioutput GP in which each output represents the DRT at a single time constant in the discretization grid. Since impedance can vary by orders of magnitude within the search space, the model is fitted to the log-scaled DRT. For computational efficiency, all outputs share the same kernel. The model is implemented in GPyTorch [42] with weakly informative hyperparameter priors that help avoid overfitting in the early stages of the experiment.

For Bayesian optimization of E_a , the log-scaled R_p is the only model output. The input features are electrode row m and column n and inverse temperature T^{-1} . Temperatures are inverted since the relation between inverse temperature and the $\ln R_p$ output is anticipated to be linear. We then use a composition kernel that is linear along the inverse temperature dimension, but uses the standard RBF covariance in the electrode row and column dimensions. The composition kernel is given by

$$k(m, n, T^{-1}) = k_1(m, n) + k_2(m, n) \times k_3(T^{-1}), \quad (A1)$$

where $k_1(m, n)$ and $k_2(m, n)$ are RBF kernels over the electrode row and column and $k_3(T^{-1})$ is a linear kernel over the inverse temperature. Intuitively, the model represents an Arrhenius description of every electrode in the library, in which the preexponential factor (represented by k_1) and activation energy (represented by k_2) are both position-dependent GPs. This physics-constrained kernel accelerates learning and increases the robustness of E_a estimation. Because E_a is proportional to the gradient of $\ln R_p$ with respect to T^{-1} , E_a also carries all the properties of a GP, such that E_a values and uncertainties can be extracted directly from the model kernel. Thus, even though the model is trained on R_p measurements, it can directly predict E_a . We contrast this with the direct approach: multiple T -dependent R_p measurements are performed for a single electrode; E_a is estimated via a linear fit of $\ln R_p$ vs T^{-1} ; and finally E_a is provided to the model. With the gradient model, E_a can be estimated without requiring multiple measurements per electrode, and the error in R_p measurements is directly accounted for, rather than being hidden in the linear fits underlying the E_a values.

3. Acquisition functions

In materials discovery, we must strike a balance between gathering novel information (exploration) and pursuing direct property improvements based on our current understanding (exploitation) [48–50]. Acquisition functions are designed to manage this trade-off: autonomous search

strategies commonly define acquisition functions that balance these considerations with a Bayesian perspective (Bayesian optimization) [40,45,49,51,52].

The surrogate model can predict property values and corresponding uncertainties for any coordinates in the search space. For any search space coordinates x_* , an acquisition function α takes the predicted property value $f(x_*)$ and variance $\sigma^2(x_*)$ at those coordinates as inputs and returns a scalar value that quantifies the priority of taking a measurement at those coordinates:

$$\alpha(x_*) = g(f(x_*), \sigma^2(x_*)). \quad (A2)$$

The coordinates at which to perform the next measurement are then selected by maximizing the acquisition function, i.e.,

$$x_{i+1} = \operatorname{argmax}_x \alpha(x). \quad (A3)$$

For active learning of the DRT, reduction of model uncertainty is the sole objective. The simplest acquisition function for this case is the pointwise variance:

$$\sigma_{PV}(x_*) = \sigma^2(x_*). \quad (A4)$$

With this acquisition function, the coordinates at which the model uncertainty is greatest will always be selected. However, coordinates at or near the edges and corners of the search space tend to remain more uncertain than those in the interior of the space, even after they have been measured. To account for this, we instead use the reduction in pointwise variance that can be expected if the test point is measured:

$$\sigma_{PVR}(x_*) = \sigma^2(x_*) - \sigma_*^2(x_*), \quad (A5)$$

where σ^2 represents the current model variance and σ_*^2 represents the expected variance after measurement at the test coordinates x_* . While simple to calculate, PVR is somewhat myopic, as it considers only how the variance changes at a single point in the search space. For a more comprehensive estimate of how a new measurement will change our state of knowledge, we also consider the IVR over the entire search space:

$$\sigma_{IVR}(x_*) = \int_x (\sigma^2(x) - \sigma_*^2(x)) dx. \quad (A6)$$

Unlike the PVR, the IVR accounts for global uncertainty reduction. Although it cannot be evaluated analytically, it can be estimated via random sampling [43].

For E_a optimization, we begin with the standard PI function:

$$\sigma_{PI}(x_*) = \Phi(y' - f(x_*), \sigma^2(x_*)), \quad (A7)$$

where Φ is the Gaussian cumulative distribution function and y' is the best (lowest) property value observed thus

far. PI estimates how likely we are to observe a new property minimum at x_* , thereby balancing exploration and exploitation [40]. However, since we are optimizing E_a but measuring R_p , use of this acquisition function directly is problematic: coordinates can retain high E_a uncertainty even after R_p measurement, which means that the standard PI function may indefinitely reselect previously measured points. Thus, we modify PI to account for information gained through R_p evaluation:

$$\sigma_{\text{GPI}}(x_*) = \Phi(y' - f(x_*), \sigma^2(x_*)) \sigma_{R_p}^2(x_*) \quad (\text{A8})$$

$$= \sigma_{\text{PI}}(x_*) \sigma_{R_p}^2(x_*), \quad (\text{A9})$$

where σ^2 and $\sigma_{R_p}^2$ are the E_a and R_p variance, respectively. As shown in Fig. S25 in Supplemental Material [34], the modified PI function accounts both for the opportunity to learn new information about R_p (and thus about E_a) and for the likelihood of reducing E_a .

We build upon the Bayesian optimization software package BoTorch [43] to implement these acquisition functions.

4. Cost-aware acquisition functions

The acquisition functions described thus far quantify the value to be gained by taking a measurement at a particular coordinate set. However, we may also consider the cost of taking a measurement. In this study, cost is simply the total time required to perform a measurement, which includes the ramping time (to adjust the controllers to the new set points), the equilibration time (to let the sample adjust to the new conditions), and the actual measurement time. We can construct a cost function $\mathcal{C}$ to describe the expected duration of evaluating a test point, which depends on both the current coordinates x_i and the test coordinates x_* :

$$\mathcal{C} = \mathcal{C}(x_*, x_i). \quad (\text{A10})$$

We can then construct a cost-aware acquisition that balances value gained with resource expenditure:

$$\alpha_{\mathcal{C}}(x_*, x_i) = \frac{\alpha(x_*)}{\mathcal{C}(x_*, x_i)}. \quad (\text{A11})$$

Because the base acquisition function is divided by the expected duration, the cost-aware acquisition function seeks to maximize the rate of utility gain per time, rather than the rate of gain per iteration.

A cost function was constructed empirically to reflect each controller's response and settling time as well as typical sample equilibration times. The direct measurement time and minimum equilibration time represent a fixed cost independent of the test coordinates, which is set to a fixed value of 60 s. Cost contribution functions for ramping time

and equilibration time are defined independently for each search space parameter, as described below.

For electrode position (row and column) and stage temperature, the cost contribution function is given by

$$c(\delta) = m(f(\delta) - f(0)) + b(1 - \phi(\delta)), \quad (\text{A12})$$

where δ represents the physical distance between electrodes or the change in stage temperature, m and b are constants, $f(x)$ is a smooth approximation of the absolute value function given by

$$f(x) = \frac{1}{\alpha} \ln(e^{-\alpha x} + e^{\alpha x}), \quad (\text{A13})$$

and $\phi(x)$ is the Gaussian RBF with length scale ρ :

$$\phi(x) = \exp\left(-\left(\frac{x}{\rho}\right)^2\right). \quad (\text{A14})$$

Equation (A12) reflects the fact that staying at the same electrode or stage temperature incurs no cost, while making a change incurs a nearly fixed overhead cost due to controller settling time (expressed by the RBF) and a variable cost that increases with the magnitude of the change (expressed by the smooth absolute value function). The values of m , b , α , and ρ were determined empirically: for electrode position, $m = 1$, $b = 5$, $\alpha = 16$, and $\rho = 0.25$; for stage temperature, $m = 1.7$, $b = 160$, $\alpha = 16$, and $\rho = 0.5$.

A different cost contribution function is defined for p_{O_2} and η to reflect nonlinear scaling of time cost with δ . Upon selection of new search space coordinates, the selected electrode begins at certain initial conditions that influence its equilibration. If the selected electrode is the same electrode that was measured in the last iteration, then the initial p_{O_2} value at the selected electrode is equal to the p_{O_2} value applied in the previous iteration. Otherwise, the initial p_{O_2} is determined by the ambient atmosphere because gas flow is controlled only at the probe tip. Thus, it is the change $\delta = \ln(p_{\text{O}_2,*} + 2 \times 10^{-4}) - \ln(p_{\text{O}_2,i} + 2 \times 10^{-4})$ in the log-transformed partial pressure, where $p_{\text{O}_2,i}$ is the relevant starting value, that determines the cost. In the case of dc voltage, the initial value faced by the electrode is always the open-circuit voltage because the measurement circuit is opened at the conclusion of each measurement. Thus, $\delta = \eta^*$ because the open-circuit voltage is approximately zero. The cost contribution is given by

$$c(\delta) = a(\ln f(b\delta) - \ln f(0)), \quad (\text{A15})$$

where $f(x)$ is the smooth absolute value function with $\alpha = 8$, and a and b are empirically determined constants. For p_{O_2} , $a = 5$ and $b = 0.5$; for η , $a = 15$ and $b = 5$. The logarithmic scaling of $f(b\delta)$ reflects declining incremental

equilibration time with increasingly large changes in p_{O_2} or η .

The individual cost contributions are combined to estimate the total cost. The change in electrode position and subsequent equilibration of the new electrode to the new p_{O_2} value occur in series, so these costs are summed. Meanwhile, the change in stage temperature occurs in parallel, so the maximum of the temperature cost contribution and the summed position and p_{O_2} costs reflects the expected cost for these changes. The equilibration of the new electrode to the selected η value begins once all other conditions are equilibrated, so this cost is added to the cost of the first set of steps. Finally, this search-parameter-dependent cost is added to the fixed cost of 60 s to obtain the overall cost. The cost contributions and overall cost function are illustrated in Fig. S3, which illustrates that changing the sample temperature incurs the largest costs due to long ramping, settling, and equilibration times.

5. Data quality validation

Data quality checks are essential in the autonomous setting. To evaluate the validity of each hybrid impedance measurement, we use the Bayesian log-likelihood of the DRT fit of the joint-domain data. The log-likelihood $\ln \mathcal{L}$ (excluding additive constants) is given by

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{i=1}^N \left(\frac{y_{i,\text{meas}} - y_{i,\text{DRT}}}{\sigma_i} \right)^2 - \sum_{i=1}^N \ln \sigma_i, \quad (\text{A16})$$

where N is the number of data points in the measurement, $y_{i,\text{meas}}$ is the measured value (voltage or impedance) at point i , $y_{i,\text{DRT}}$ is the value given by the DRT fit, and σ_i is the square root of the estimated error variance. Larger log-likelihood values indicate better agreement between the model and the experiment. A normalized log-likelihood value is obtained by dividing the expression in Eq. (A16) by the number of data points N .

To identify invalid measurements (due to poor electrical contact, instability, or other measurement issues), we use a Bayesian outlier test. We assume that the distribution of log-likelihood values comprises two groups: inliers (normal, valid measurements) and outliers (abnormal, likely invalid measurements). We assume that each group is normally distributed with its own mean and variance:

$$\ln \mathcal{L}_{\text{in}} \sim \mathcal{N}(\mu_{\text{in}}, \sigma_{\text{in}}^2), \quad (\text{A17})$$

$$\ln \mathcal{L}_{\text{out}} \sim \mathcal{N}(\mu_{\text{out}}, \sigma_{\text{out}}^2). \quad (\text{A18})$$

Although outliers may follow a non-normal distribution, we treat them with a normal distribution for simplicity, as it is not obvious *a priori* which kind of distribution they will follow. Because the inlier and outlier distributions are distinguished primarily by their spreads, the outlier test is

not overly sensitive to the choice of outlier distribution. We use a Bernoulli distribution to define the prior probability that any measurement is an outlier:

$$k \sim B(p_{\text{out}}), \quad (\text{A19})$$

where $k = 1$ (with prior probability p_{out}) indicates the outlier group and $k = 0$ (with prior probability $1 - p_{\text{out}}$) indicates the inlier group. We can then define the posterior probability that any measurement is part of the outlier group as

$$P(k = 1 \mid \ln \mathcal{L}) = \frac{f(\ln \mathcal{L} \mid k = 1)p_{\text{out}}}{f(\ln \mathcal{L} \mid k = 0)(1 - p_{\text{out}}) + f(\ln \mathcal{L} \mid k = 1)p_{\text{out}}}, \quad (\text{A20})$$

where $f(\ln \mathcal{L} \mid k = 0)$ and $f(\ln \mathcal{L} \mid k = 1)$ are the $\ln \mathcal{L}$ probability density functions for the inlier and outlier groups:

$$f(\ln \mathcal{L} \mid k = 0) = \frac{1}{\sqrt{2\pi\sigma_{\text{in}}^2}} \exp\left(-\frac{1}{2} \left(\frac{\ln \mathcal{L} - \mu_{\text{in}}}{\sigma_{\text{in}}}\right)^2\right), \quad (\text{A21})$$

$$f(\ln \mathcal{L} \mid k = 1) = \frac{1}{\sqrt{2\pi\sigma_{\text{out}}^2}} \exp\left(-\frac{1}{2} \left(\frac{\ln \mathcal{L} - \mu_{\text{out}}}{\sigma_{\text{out}}}\right)^2\right). \quad (\text{A22})$$

To use the above equations, we first estimate the parameters of the inlier distribution by calculating the median and interquartile range (r_{IQ}) of the observed log-likelihood values, which should be outlier-robust metrics. We set μ_{in} to the median and $\sigma_{\text{in}} = r_{\text{IQ}}/1.349$, the values expected for a normally distributed variable. We then use the sample mean and variance of any points lying outside the window $[\mu_{\text{in}} - \sigma_{\text{in}}, \mu_{\text{in}} + \sigma_{\text{in}}]$ to estimate μ_{out} and σ_{out}^2 for the hypothetical outlier group. The estimated outlier standard deviation is forced to be at least twice the inlier standard deviation. We set the Bernoulli prior $p_{\text{out}} = 0.01$. Equation (A20) is then evaluated for all measurements, and any measurements for which $P(k = 1 \mid \ln \mathcal{L}) > 0.9$ are flagged as invalid and removed from the model training set. This flagging procedure is initiated once 20 measurements have been collected (to ensure sufficient sample size) and is repeated at each subsequent iteration of the experiment.

-
- [1] J. H. Montoya, K. T. Winther, R. A. Flores, T. Bligaard, J. S. Hummelshøj, and M. Aykol, Autonomous intelligent agents for accelerated materials discovery, *Chem. Sci.* **11**, 8517 (2020).

- [2] S. G. Louie, Y. H. Chan, F. H. da Jornada, Z. Li, and D. Y. Qiu, Discovering and understanding materials through computation, *Nat. Mater.* **20**, 728 (2021).
- [3] E. Annevelink, R. Kurchin, E. Muckley, L. Kavalsky, V. I. Hegde, V. Sulzer, S. Zhu, J. Pu, D. Farina, M. Johnson, D. Gandhi, A. Dave, H. Lin, A. Edelman, B. Ramsundar, J. Saal, C. Rackauckas, V. Shah, B. Meredig, and V. Viswanathan, AutoMat: Automated materials discovery for electrochemical systems, *MRS Bull.* **47**, 1036 (2022).
- [4] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, Scaling deep learning for materials discovery, *Nature* **624**, 80 (2023).
- [5] A. Benayad, D. Diddens, A. Heuer, A. N. Krishnamoorthy, M. Maiti, F. L. Cras, M. Legallais, F. Rahmanian, Y. Shin, H. Stein, M. Winter, C. Wölke, P. Yan, and I. Cekic-Laskovic, High-throughput experimentation and computational freeway lanes for accelerated battery electrolyte and interface development research, *Adv. Energy Mater.* **12**, 2102678 (2022).
- [6] H. S. Stein, A. Sanin, F. Rahmanian, B. Zhang, M. Vogler, J. K. Flowers, L. Fischer, S. Fuchs, N. Choudhary, and L. Schroeder, From materials discovery to system optimization by integrating combinatorial electrochemistry and data science, *Curr. Opin. Electrochem.* **35**, 101053 (2022).
- [7] J. M. Gregoire, L. Zhou, and J. A. Haber, Combinatorial synthesis for AI-driven materials discovery, *Nat. Synth.* **2**, 493 (2023).
- [8] A. M. Lunt, H. Fakhruddin, G. Pizzuto, L. Longley, A. White, N. Rankin, R. Clowes, B. Alston, L. Gigli, G. M. Day, A. I. Cooper, and S. Y. Chong, Modular, multi-robot integration of laboratories: An autonomous workflow for solid-state chemistry, *Chem. Sci.* **15**, 2456 (2024).
- [9] M. Abolhasani and E. Kumacheva, The rise of self-driving labs in chemical and materials sciences, *Nat. Synth.* **2**, 483 (2023).
- [10] G. Tom, S. P. Schmid, S. G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo, S. Pablo-García, E. M. Rajaonson, M. Skreta, N. Yoshikawa, S. Corapi, G. D. Akkoc, F. Strieth-Kalthoff, M. Seifrid, and A. Aspuru-Guzik, Self-driving laboratories for chemistry and materials science, *Chem. Rev.* **124**, 9633 (2024).
- [11] B. P. MacLeod, F. G. Parlane, C. C. Rupnow, K. E. Dettelbach, M. S. Elliott, T. D. Morrissey, T. H. Haley, O. Proskurin, M. B. Rooney, N. Taherimakhsoosi, D. J. Dvorak, H. N. Chiu, C. E. Waizenegger, K. Ocean, M. Mokhtari, and C. P. Berlinguette, A self-driving laboratory advances the Pareto front for material properties, *Nat. Commun.* **13**, 1 (2022).
- [12] B. P. MacLeod, F. G. Parlane, T. D. Morrissey, F. Häse, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. Yunker, M. B. Rooney, J. R. Deeth, V. Lai, G. J. Ng, H. Situ, R. H. Zhang, M. S. Elliott, T. H. Haley, D. J. Dvorak, A. Aspuru-Guzik, J. E. Hein, and C. P. Berlinguette, Self-driving laboratory for accelerated discovery of thin-film materials, *Sci. Adv.* **6**, 1 (2020).
- [13] N. J. Szymanski, P. Nevatia, C. J. Bartel, Y. Zeng, and G. Ceder, Autonomous and dynamic precursor selection for solid-state materials synthesis, *Nat. Commun.* **14**, 6956 (2023).
- [14] N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng, and G. Ceder, An autonomous laboratory for the accelerated synthesis of inorganic materials, *Nature* **624**, 86 (2023).
- [15] F. Strieth-Kalthoff *et al.*, Delocalized, asynchronous, closed-loop discovery of organic laser emitters, *Science* **384**, eadk9227 (2024).
- [16] M. Vogler, S. K. Steensen, F. F. Ramírez, L. Merker, J. Busk, J. M. Carlsson, L. H. Rieger, B. Zhang, F. Liot, G. Pizzi, *et al.*, Autonomous battery optimization by deploying distributed experiments and simulations, *Adv. Energy Mater.* **14**, 2403263 (2024).
- [17] N. J. Szymanski, Y. Zeng, H. Huo, C. J. Bartel, H. Kim, and G. Ceder, Toward autonomous design and synthesis of novel inorganic materials, *Mater. Horiz.* **8**, 2169 (2021).
- [18] H. Jorress, M. L. Green, I. Takeuchi, and J. R. Hattrick-Simpers, *Applications of High Throughput (Combinatorial) Methodologies to Electronic, Magnetic, Structural, and Energy-Related Materials*, edited by F. G. Caballero (Elsevier, Oxford, 2021), pp. 353–371.
- [19] M. C. Papac, J. Huang, A. Zakutayev, and R. O’Hayre, Combinatorial impedance spectroscopy with Bayesian analysis for triple ionic-electronic conducting perovskites, *J. Mater. Chem. A* **11**, 5267 (2023).
- [20] R. Shimizu, S. Kobayashi, Y. Watanabe, Y. Ando, and T. Hitosugi, Autonomous materials synthesis by machine learning and robotics, *APL Mater.* **8**, 111110 (2020).
- [21] I. Ohkubo, Z. Hou, J. N. Lee, T. Aizawa, M. Lippmaa, T. Chikyow, K. Tsuda, and T. Mori, Realization of closed-loop optimization of epitaxial titanium nitride thin-film growth via machine learning, *Mater. Today Phys.* **16**, 100296 (2021).
- [22] Y. K. Wakabayashi, T. Otsuka, Y. Krockenberger, H. Sawada, Y. Taniyasu, and H. Yamamoto, Bayesian optimization with experimental failure for high-throughput materials growth, *npj Comput. Mater.* **8**, 1 (2022).
- [23] D. M. Fébba, K. R. Talley, K. Johnson, S. Schaefer, S. R. Bauers, J. S. Mangum, R. W. Smaha, and A. Zakutayev, Autonomous sputter synthesis of thin film nitrides with composition controlled by Bayesian optimization of optical plasma emission, *APL Mater.* **11**, 071119 (2023).
- [24] F. Ciucci, Modeling electrochemical impedance spectroscopy, *Curr. Opin. Electrochem.* **13**, 132 (2019).
- [25] S. Wang, J. Zhang, O. Gharbi, V. Vivier, M. Gao, and M. E. Orazem, Electrochemical impedance spectroscopy, *Nat. Rev. Methods Primers* **1** (2021).
- [26] Q. Liang, A. E. Gongora, Z. Ren, A. Tiitonen, Z. Liu, S. Sun, J. R. Deneault, D. Bash, F. Mekki-Berrada, S. A. Khan, *et al.*, Benchmarking the performance of Bayesian optimization across multiple experimental materials science domains, *npj Comput. Mater.* **7**, 188 (2021).
- [27] M. C. Papac, K. R. Talley, R. O’Hayre, and A. Zakutayev, Instrument for spatially resolved, temperature-dependent electrochemical impedance spectroscopy of thin films under locally controlled atmosphere, *Rev. Sci. Instrum.* **92**, 065105 (2021).
- [28] C. Duan, J. Tong, M. Shang, S. Nikodemski, M. Sanders, S. Ricote, A. Almansoori, and R. O’Hayre, Readily processed

- protonic ceramic fuel cells with high performance at low temperatures, *Science* **349**, 1321 (2015).
- [29] C. Duan, R. J. Kee, H. Zhu, C. Karakaya, Y. Chen, S. Ricote, A. Jarry, E. J. Crumlin, D. Hook, R. Braun, N. P. Sullivan, and R. O'Hayre, Highly durable, coking and sulfur tolerant, fuel-flexible protonic ceramic fuel cells, *Nature* **557**, 217 (2018).
- [30] J. D. Huang, C. Meisel, N. P. Sullivan, A. Zakutayev, and R. O'Hayre, Rapid mapping of electrochemical processes in energy-conversion devices, *Joule* **8**, 2049 (2024).
- [31] J. Huang, M. Papac, and R. O'Hayre, Towards robust autonomous impedance spectroscopy analysis: A calibrated hierarchical Bayesian approach for electrochemical impedance spectroscopy (EIS) inversion, *Electrochim. Acta* **367**, 137493 (2021).
- [32] J. D. Huang, N. P. Sullivan, A. Zakutayev, and R. O'Hayre, How reliable is distribution of relaxation times (DRT) analysis? A dual regression-classification perspective on DRT estimation, interpretation, and accuracy, *Electrochim. Acta* **443**, 141879 (2023).
- [33] Y. Shin, Y.-d. Kim, M. Sanders, S. P. Harvey, M. Walker, and R. O'Hayre, Tuning the Co/Fe ratio in $\text{BaCo}_x\text{Fe}_{0.8-x}\text{Zr}_{0.1}\text{Y}_{0.1}\text{O}_{3-\delta}$, a promising triple ionic and electronic conducting oxide, to boost electrolysis and fuel cell performance, *J. Mater. Chem. A* **10**, 24839 (2022).
- [34] See Supplemental Material <http://link.aps.org/supplemental/10.1103/nkhy-y2nz>, in which Refs. [53–57] are cited, for additional method details and supporting figures.
- [35] J. Xia, C. Wang, X. Wang, L. Bi, and Y. Zhang, A perspective on DRT applications for the analysis of solid oxide cell electrodes, *Electrochim. Acta* **349**, 136328 (2020).
- [36] C. Plank, T. R  ther, L. Jahn, M. Schamel, J. P. Schmidt, F. Ciucci, and M. A. Danzer, A review on the distribution of relaxation times analysis: A powerful tool for process identification of electrochemical systems, *J. Power Sources* **594**, 233845 (2024).
- [37] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2009).
- [38] D. Poetzsch, R. Merkle, and J. Maier, Oxygen reduction at dense thin-film microelectrodes on a proton-conducting electrolyte: I. Considerations on reaction mechanism and electronic leakage effects, *J. Electrochem. Soc.* **162**, F939 (2015).
- [39] S. Kullback and R. A. Leibler, On information and sufficiency, *Ann. Math. Stat.* **22**, 79 (1951).
- [40] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, Taking the human out of the loop: A review of Bayesian optimization, *Proc. IEEE* **104**, 148 (2016).
- [41] J. Huang, M. Papac, D. F  bba, R. O'Hayre, and A. Zakutayev, Data and code repository for: A benchtop autonomous electrochemical characterization system for combinatorial thin-film solid oxide electrodes, Zenodo, <https://doi.org/10.5281/zenodo.18679273> (2026).
- [42] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson, in *Advances in Neural Information Processing Systems* (Neural Information Processing Systems Foundation, Inc. (NeurIPS), Montreal, Canada, 2018).
- [43] M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, in *Advances in Neural Information Processing Systems* (Neural Information Processing Systems Foundation, Inc. (NeurIPS), Vancouver, Canada, 2020), Vol. 33.
- [44] M. Olson *et al.*, in *AutoML 2025 ABCD Track* (Open Review, New York, 2025).
- [45] P. I. Frazier and J. Wang, in *Information Science for Materials Discovery and Design. Springer Series in Materials Science*, edited by T. Lookman, F. Alexander and K. Rajan (Springer, Cham, 2015), Vol. 225.
- [46] M. M. Noack, G. S. Doerk, R. Li, J. K. Streit, R. A. Vaia, K. G. Yager, and M. Fukuto, Autonomous materials discovery driven by Gaussian process regression with inhomogeneous measurement noise and anisotropic kernels, *Sci. Rep.* **10**, 1 (2020).
- [47] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (The MIT Press, Cambridge, Massachusetts, 2006).
- [48] J. Ling, M. Hutchinson, E. Antono, S. Paradiso, and B. Meredig, High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates, *Integr. Mater. Manuf. Innovation* **6**, 207 (2017).
- [49] P. V. Balachandran, Adaptive machine learning for efficient materials design, *MRS Bull.* **45**, 579 (2020).
- [50] B. Rohr, H. S. Stein, D. Guevarra, Y. Wang, J. A. Haber, M. Aykol, S. K. Suram, and J. Gregoire, Benchmarking the acceleration of materials discovery by sequential learning, *Chem. Sci.* **11**, 2696 (2020).
- [51] M. Malu, G. Dasarathy, and A. Spanias, in *IISA 2021 – 12th International Conference on Information, Intelligence, Systems and Applications* (IEEE, Chania Crete, Greece, 2021).
- [52] T. Ueno, T. D. Rhone, Z. Hou, T. Mizoguchi, and K. Tsuda, COMBO: An efficient Bayesian optimization library for materials science, *Mater. Discover* **4**, 18 (2016).
- [53] H. G. Bohn and T. Schober, Electrical conductivity of the high-temperature proton conductor $\text{BaZr}_{0.9}\text{Y}_{0.1}\text{O}_{2.95}$, *J. Am. Ceram. Soc.* **83**, 768 (2000).
- [54] H. Zhu, S. Ricote, C. Duan, R. P. O'Hayre, D. S. Tsvetkov, and R. J. Kee, Defect incorporation and transport within dense $\text{BaZr}_{0.8}\text{Y}_{0.2}\text{O}_{3-\delta}$ (BZY20) proton-conducting membranes, *J. Electrochem. Soc.* **165**, F581 (2018).
- [55] K. R. Talley, S. R. Bauers, C. L. Melamed, M. C. Papac, K. N. Heinselman, I. Khan, D. M. Roberts, V. Jacobson, A. Mis, G. L. Brennecke, J. D. Perkins, and A. Zakutayev, COMBIgor: Data-analysis package for combinatorial materials science, *ACS Comb. Sci.* **21**, 537 (2019).
- [56] M. C. Henstridge, E. Laborda, and R. G. Compton, Asymmetric Marcus–Hush model of electron transfer kinetics: Application to the voltammetry of surface-bound redox systems, *J. Electroanal. Chem.* **674**, 90 (2012).
- [57] Y. Zeng, P. Bai, R. B. Smith, and M. Z. Bazant, Simple formula for asymmetric Marcus–Hush kinetics, *J. Electroanal. Chem.* **748**, 52 (2015).