Adaptive learning-driven high-throughput synthesis of oxygen reduction reaction Fe-N-C electrocatalysts

de Melo Kort-Kamp, Wilton Junior
Ferrandon, Magali
Wang, Xiaoping
Park, Jae
Malla, Rajesh
Ahmed, Towfiq
Holby, Edward F.
Myers, Deborah
Zelenay, Piotr

To be published in: Journal of Power Sources

DOI to publisher's version: 10.1016/j.jpowsour.2022.232583

Adaptive Learning-Driven High-throughput Synthesis of Oxygen Reduction Reaction Fe-N-C Electrocatalysts

Wilton J. M. Kort-Kamp[a], Magali Ferrandon[b], Xiaoping Wang[b], Jae Hyung Park[b], Rajesh K. Malla[c], Towfiq Ahmed[d], Edward F. Holby[e], Deborah J. Myers[b], Piotr Zelenay[f]

[a] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, United States*

[b] *Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, IL 60439, United States*

[c] *Condensed Matter Physics and Materials Science Division, Brookhaven National Laboratory, Upton, NY 11973, United States*

[d] *National Security Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, United States*

[e] *Sigma Division, Los Alamos National Laboratory, Los Alamos, NM 87545, United States*

[f] *Materials Physics and Applications Division, Los Alamos National Laboratory, Los Alamos, NM 87545, United States*

* Corresponding author: kortkamp@lanl.gov

Abstract

Reducing human reliance on inefficient energy systems and fossil fuels has become more urgent due to the consequences of global climate change. However, traditional trial-and-error approaches have hampered our ability to accelerate the discovery and implementation of functional materials for efficient energy conversion devices, such as polymer electrolyte fuel cells (PEFCs). To address this, we develop an adaptive learning framework that integrates machine learning and state-of-the-art capabilities in high-throughput synthesis to achieve expedited optimization of iron-nitrogen-carbon PEFC oxygen reduction reaction (ORR) electrocatalysts. We use statistical inference, uncertainty quantification, and global optimization to build a computational design-of-experiment tool that identifies the optimum compositions to be investigated next to reduce the demands placed on experimental materials discovery. We benchmark the ability of the proposed strategy to discover optimum catalyst synthesis conditions in a six-dimensional search space when starting with a thirty-six-sample database. By following the adaptive learning strategy, we synthesize fourteen new catalysts from approximately ten billion unique compositions and discover four catalysts that outperform all original samples. The best machine learning-optimized catalyst is 33% more active than the highest-performing one in the initial database, showing an ORR activity seven times larger than those typically reported for the same class of materials.

Keywords

machine learning; uncertainty quantification; high-throughput synthesis; iron-nitrogen-carbon electrocatalysts; oxygen reduction reaction; hydrogen fuel cells

Highlights

- Optimized Fe-N-C oxygen reduction reaction catalysts via adaptive learning
- Elucidated the role of high-throughput synthesis conditions on catalyst activity
- Developed a computational design-of-experiment tool to pick catalysts to test next
- Discovered four catalysts that outperform the best materials in the initial dataset
- Demonstrated an activity increase of ~33% relative to the previous best sample

1. Introduction

High-performance and cost-competitive hydrogen fuel cells comprised of earth-abundant materials have been a sought-after goal that would have a transformative effect on renewable energy technologies [1, 2]. In particular, the design and experimental realization of electrocatalysts addressing the sluggish kinetics of the oxygen reduction reaction (ORR) and resulting high ORR overpotentials in polymer electrolyte fuel cells (PEFCs) [3-6] could enable large-scale deployment of these electrochemical devices with major economic and energy security benefits. The best ORR catalysts are based on platinum, which is one of the main factors limiting large-scale application of the technology in automotive transportation due to the high cost and geographic scarcity of platinum [7, 8]. The development of alternative catalysts derived from earth-abundant materials, targeting activity and stability comparable to or higher than those of platinum, has thus been an active research area for several decades. Although progress in developing novel platinum group metal-free (PGM-free) catalysts with enhanced ORR performance [8-16] has been made, significant improvements to ORR catalyst activity and performance durability are still needed to make them viable for real-world applications.

PGM-free catalysts with the highest ORR activity have been derived from iron precursors and metal-organic frameworks, as well as carbon-nitrogen-containing polymers [15,17-20]. While representing substantial progress in the field, the ORR turnover frequency and volumetric active site density of these materials are still approximately one order of magnitude lower than those of the state-of-the-art platinum alloy nanoparticle catalysts [13,16]. For the general class of pyrolyzed iron-nitrogen-carbon PGM-free materials, it has been determined that variables such as the iron loading, the type of the carbon and nitrogen precursors, their relative concentrations, as well as the temperature, heating and cooling rates, pyrolysis time, and pyrolysis atmosphere are critical in

determining the activity and stability of the electrocatalysts [9,14,17, 21-23]. The experimental effort required to explore the effect of all these variables on the catalyst properties is an overwhelmingly time-consuming process. Most approaches seeking to develop optimum synthesis paths have been guided by designer's intuition, with only a small fraction of the composition and temperature search space explored to date for this broad class of materials [24]. However, the interplay between the multiple degrees of freedom in typical synthesis methods is often complex and the combinatorial possibilities are too large to be practically handled via trial-and-error approaches, resulting in suboptimal outcomes.

Machine learning (ML) has become a promising tool for materials science [25, 26] and has been a trending approach to tackle challenging problems in renewable energy [27-29]. It has enabled the development and implementation of elaborate numerical models that surrogate multidimensional data for expedited discovery, optimization, and property prediction of electrocatalysts and photocatalysts [30-33]. Recent works have also leveraged the capabilities of ML algorithms to guide experimental efforts in batteries [34, 35], solid oxide fuel cells [36], and PEFCs electrode development [37]. The focus of these efforts has often been on feature selection and dimensionality reduction of the search space or the use of regression techniques to predict material properties (dependent outcomes) given a set of parameter values describing the system (independent variables). These investigations routinely have little or no iterative experimental feedback, are susceptible to large uncertainty due to the vastness of the search space, and suffer from the usually limited availability of data, leading to a poor generalization of the model. Moreover, exclusive exploitation of model outputs may result in suboptimal learning due to existing local maxima, which prevent the exploration of regions of the search space with greater uncertainty. These challenges can be overcome through adaptive learning frameworks:

computational methods that leverage machine learning surrogate models and their uncertainties to balance the trade-offs between testing materials likely to perform better (exploitation) versus probing materials with large uncertainty (exploration), but which could improve the quality of the models in the long term. Adaptive learning algorithms minimize the number of measurements needed to achieve an improved property outcome via an iterative positive feedback loop that provides a robust guide for optimally selecting the next set of experiments. These computational methods have been successfully demonstrated, among others, in computer science [38], solid-state physics [39], and electrochemical reduction of CO_2 [40]. Optimizing the synthesis of PGM-free ORR electrocatalysts via adaptive design remains largely uncharted to date.

In this paper, we introduce an adaptive learning framework tailored to optimizing the synthesis conditions for Fe-N-C-based ORR catalysts with enhanced activity and durability, as measured via the rotating ring disk electrode (RRDE) technique. Inspired by the method developed by Proietti *et al.* [17], we synthesize nitrogen-coordinated iron-doped carbon electrocatalysts using physical mixing, through ball milling, of a zeolitic imidazolate framework (ZIF-8), 1,10-phenanthroline, and iron acetate. Ball milling is a process in which the moving balls apply their kinetic energy to the milled material, breaking chemical bonds, triggering solid-solid reactions, causing changes in the materials morphology, and possibly forming structural defects that can increase the density of active catalytic sites [41, 42]. Unlike liquid phase synthesis of doped ZIFs, high-energy ball milling followed by controlled thermal annealing is an easy and scalable method in which chemical compositions and concentrations can be easily varied [43]. Our experiments leverage our state-of-the-art robotic platform and the high-energy ball milling apparatus in Argonne National Laboratory's High-Throughput Research Facility to synthesize numerous ORR electrocatalyst compositions using a facile synthetic procedure and commercially available precursors. We

develop and apply our adaptive learning approach to optimize six independently controlled synthesis variables, including the iron weight loading, phenanthroline-to-ZIF-8 ratio, pyrolysis temperature, heating rate, pyrolysis time, and fast versus slow cooling rates. Even in this restricted search space, there are about 10^{10} combinations of synthesis conditions (see Section 2.3.1 for details), making it impossible to experimentally explore all cases. However, our high-throughput experimental capabilities have allowed for benchmarking our adaptive design framework, which guided the synthesis of fourteen new Fe-N-C-based catalysts. We demonstrate that four of the machine learning-selected materials have ORR activity higher than that of the thirty-six samples in the original dataset used for model development. Our approach reduces the number of costly, time-consuming experiments and paves the way for integration of data science with high-throughput experiments for accelerated discovery of highly active PGM-free electrocatalysts for a variety of electrochemical reactions, beyond oxygen reduction considered here.

2. Methods

2.1. Adaptive Learning Design Framework

A schematic representation of the adaptive learning framework that we developed for discovering optimum high-throughput synthesis conditions for Fe-N-C-based ORR electrocatalysts is presented in **Fig. 1**. The adaptive learning process involves four critical steps synergistically integrated to iteratively guide the next set of experiments based on the knowledge gained from existing measurements. The first key element in our design framework is the generation of a database with various materials (compositions) labeled by a set of descriptors and associated outcomes that have already been measured. In our specific problem, we index the catalyst samples according to the features independently controlled during the synthesis process and use the ORR activity in the first and second RRDE cycles of staircase voltammetry as the

properties of interest (see the Experimental section below for details of sample characterization). This database is then used to train and cross-validate machine learning regression models that learn the relationships between input parameters and output outcomes. Due to the sparse availability of initial data, statistical methods are integrated with ML algorithms to quantify the uncertainty associated with the model predictions. A design-of-experiment (DoE) tool is then developed in the third step of the adaptive design framework by combining machine learning and uncertainty quantification with global optimization approaches. This tool leverages model predictions and their errors to conduct a search in the space of unexplored materials (not yet synthesized and experimentally characterized) and chooses the best sample to be tested in the next round of experiments by weighing the benefits between exploitation and exploration. Finally, the new suggested material is synthesized and its properties are measured to compare against computational results and to augment the database for improvement of the machine learning regression models in subsequent iterations of the design loop.

2.2. Experimental

2.2.1. Materials

ZIF-8 (Sigma Aldrich), 1,10-phenanthroline (Sigma-Aldrich), and iron (II) acetate (99.99%, Sigma Aldrich) were dried under vacuum for ~12 h at 80 °C before use. All gases used during the synthesis are ultra-high purity grade (Airgas).

2.2.2. High-Throughput Catalyst Synthesis and Sample Characterization

An automated synthesis platform (Junior, Unchained Labs Inc.) in Argonne National Laboratory's High-throughput Research Facility [44], as shown in **Fig. 2a**, was used for dispensing the catalyst solid precursor materials through disposable shaker vials and weighed using an

automated balance (± 0.2 mg accuracy). Synthesis protocols were designed in Library Studio and Automation Studio (LEA software) was used to run the protocols. The ZIF-8, phenanthroline, and iron acetate were mixed and placed into a 50 mL zirconia jar with 34 zirconia beads (3 mm diameter) at a ratio of precursors:beads = 1:10 and milled at 400 rpm for 2 h in a high-throughput planetary ball mill (PM 400, Retsch), schematically shown in **Fig. 2b**. The precursor mixtures were then heated in a flowing argon atmosphere in a quartz tube in a clamshell furnace at various rates to various pyrolysis temperatures and held at those temperatures for varying lengths of time. The samples were then cooled to 900 °C in flowing argon and exposed to a flowing pure ammonia atmosphere for 5 min before cooling down to room temperature in flowing argon either naturally with the furnace closed, or rapidly by opening the furnace. The temperature profiles during the cooling period for both cases can be found in **Fig. S1** in the Supporting Material.

The ORR half-wave potential ($E_{½}$), the mass activity, and hydrogen peroxide (H_2O_2) yield were evaluated for each catalyst by using a rotating ring (Pt)-disk (glassy carbon, GC) electrode (E7R9 ThinGap RRDE Tips PTFE, Pine Research) and a CHI potentiostat (CHI 760C, CH Instruments, Inc.). These measurements utilized a three-electrode electrochemical cell, room-temperature oxygen-saturated 0.5 M H_2SO_4 electrolyte, a catalyst-ionomer layer deposited on the GC of the RRDE as the working electrode, a graphite rod counter electrode, and a Hg/Hg_2SO_4 reference electrode. While less sensitive than hydrogen-air fuel cell performance, where current density especially at voltages <0.7 V can be strongly affected by the transport of reactants to the active sites, ORR activities measured by RRDE can be influenced by the catalyst ink composition [45]. Here, we prepared all samples using the same catalyst-ionomer ink by measuring 5 mg of catalyst powder, 20 μL of Nafion™ dispersion (5 wt% Nafion™ D521, Ion Power, Inc.), and 0.5 mL of isopropanol into a glass vial, followed by sonication in an ultrasonic ice water bath for 30

min. Following sonication, the desired amount of catalyst ink was then deposited onto the GC disk of the RRDE in four aliquots, resulting in a catalyst loading of 0.6 mg/cm^2. It was observed that the Fe-N-C catalysts can undergo degradation during the various stages, from synthesis to ORR activity testing, specifically, during aging of the catalyst powder and catalyst-ionomer ink [46], as well as during electrochemical testing. To minimize the impact of catalyst degradation on the electrochemical results, the elapsed time from catalyst synthesis to electrochemical testing was carefully controlled for the different catalysts, as was the sonication time, the time interval from ink sonication to ink deposition to electrochemical testing. Before use, the potential of the Hg/Hg_2SO_4 reference electrode was calibrated versus a reversible hydrogen electrode (RHE) using a Pt coil electrode immersed in hydrogen-saturated 0.5 M H_2SO_4 electrolyte. The dried catalyst layer was preconditioned using cyclic voltammetry scans between 0 and 1.0 V vs. RHE for 20 cycles at a scan rate of 100 mV/s to remove impurities, ensure fully wetting of the catalyst layer, and stable cyclic voltammograms (CVs) before the steady-state ORR polarization curves were recorded in staircase voltammetry scans from 1.0 to 0 V vs. RHE with a potential step of 20 mV and a step period of 20 sec with iR correction in oxygen-saturated 0.5 M H_2SO_4. The rotation rate of the RRDE was controlled to 900 rpm using a Pine rotator. The ring electrode collection efficiency (N) in RRDE experiments was found to be 0.379, as determined from the ratio of the ring and disk currents of the $Fe(CN)_6^{4-/3-}$ redox couple in 1 M NaOH solution containing 10 mM $K_3[Fe(CN)_6]$. The ring potential was set to 1.2 V vs. RHE. [47, 48]. The ORR mass activities were obtained by calculating the kinetic current at 0.8 V vs RHE (using the Levich-Koutecky equation) and dividing by the catalyst mass deposited on the GC disk electrode. The H_2O_2 yields at various disk potentials were calculated from the measured ring and disk currents and collection efficiency.

2.3. Machine Learning and Statistical Inference

2.3.1. Machine Learning Descriptors and Materials Database

As depicted in **Fig. 2c**, each Fe-N-C-based electrocatalyst is uniquely indexed in terms of six independently controlled features that provide a coarse-grained and relatively simple description of the catalyst synthesis. A reduced number of descriptors could be selected based on prior experience working with this class of electrocatalysts, decreasing the dimensionality of the search space at the sacrifice of potentially missing high-performance compositions not previously investigated in the literature. Here, we decided to operate in a higher dimensional synthesis space despite increasing the computational cost. A larger number of descriptors accounting for, e.g., the composition of the catalyst ink could also have been considered. However, that would result in an increase of the already large search space (see below), adding significant complexity to the optimization problem. Hence, we evaluated the ORR activity of all catalyst samples using the same ink composition and demonstrated a proof-of-concept of the developed adaptive learning strategy considering the independent parameters discussed in **Fig. 2c**. In addition to the input parameters, we use the ORR mass activity measured in the first two RRDE scans as the dependent properties. The main target of the machine learning models is to infer a correlation map between the outputs for catalyst activity in selected RRDE cycles and the input descriptors.

Thirty-six samples were synthesized and characterized as discussed in the previous section to form the initial database (training set) in the adaptive design loop (**Tables S1** and **S2**). They were picked from pre-selected discretized values of the synthesis condition descriptors, randomly distributed over a quinary search space (Fe loading, phenanthroline-to-ZIF ratio, pyrolysis temperature, heating rate, and pyrolysis time), with half of the original dataset comprising ORR catalysts cooled down to room temperature while keeping the furnace closed or open following

the ammonia treatment at 900 ºC. **Table 1** shows the range of values considered in this work for each of the parameters. Theoretically, these five descriptors could be varied continuously, resulting in an infinite number of possible combinations. In practice, the number of unique synthesis conditions depends on how one partitions the search space. Note, however, that the maximum number of combinations that would result in unique catalysts can be estimated by considering the ranges in **Table 1** together with the uncertainty in the input parameters. Indeed, the uncertainty is the minimum step-size needed to distinguish two values of a given descriptor and, hence, serves as a metric of how finely the search space can be meshed in each dimension. In our experiments, the uncertainty is ~0.05% for the Fe weight loading, 0.002% for the phenanthroline-to-ZIF ratio, 5 ºC for the pyrolysis temperature, 5 minutes for the pyrolysis time, and 0.2 ºC/min for the heating rate. Consequently, the maximum number of potential high-throughput synthesis pathways is on the order of 10^{10} (ten billion) for each of the two cooling rates considered. Even in a more conservative scenario, e.g., by assuming a coarser discretization of the search space with uncertainties ten times larger than those used above for each input descriptor, there would still be more than 10^5 (one hundred thousand) allowed combinations, making exploration of the entire electrochemical search space unamenable to experiments.

Table 1

Independently controlled parameters in the high-throughput synthesis of Fe-N-C ORR electrocatalysts and their ranges of variation used for training machine learning surrogate models.

Fe weight loading	0.80-2.80 %	**Pyrolysis time**	0.0-3.0 h
Phen-to-ZIF-8 ratio	0.00-10.00 %	**Heating rate**	5.0-20.0 ºC/min
Pyrolysis temperature	950-1050 ºC	**Cooling rate**	Furnace closed or open

2.3.2. Machine Learning Regressors and Uncertainty Quantification

In the second step of our adaptive design framework, we trained three machine learning algorithms on the initial datasets with eighteen samples, each corresponding to the two possible furnace states (closed or open) during the cooling stage. The regression algorithms considered in this study are Gaussian Process Regression (GPR), Random Forest (RF), and Artificial Neural Network (ANN), and have been implemented in python via the SCIKIT-LEARN [49] package. Given our small database, we only split the initial data in an 80-to-20 training-to-cross-validation size ratio, and use the new samples synthesized and characterized via the adaptive learning framework as testing sets for probing the quality of the developed ML models. Hyperparameters in the GPR algorithm are automatically optimized during the training process and a superposition of the radial basis functions kernel and the white kernel was considered to specify the covariance function of the regressor. We optimized the RF and ANN algorithms via a comprehensive grid search over the hyperparameters space using ten random shuffle splits of the initial data while targeting minimization of the root-mean-squared error (RMSE) cost function.

The GPR model is very attractive for integration into adaptive learning approaches as it is a Bayesian-based approach that naturally provides a mean value (μ) of the desired output property including an uncertainty estimate (σ). On the other hand, the RF and ANN models do not generally estimate uncertainties. In these latter cases, we perform uncertainty quantification via bootstrap statistical sampling [50]. This method generates new training sets (bootstraps) by randomly selecting samples from the original database allowing for replacement, i.e., the same catalyst can be picked multiple times. We generated 1,000 bootstraps of the original data for each state of the furnace during the cooling process. We then train the RF and ANN algorithms in each bootstrap and use the predictions from the 1,000 ML models to compute the average value of the output

ORR activity. The standard deviation from predictions of bootstraps-trained ML models is also used to estimate the uncertainty for each regression algorithm. Importantly, the use of statistical bootstraps significantly reduces the chances that the ML-trained models overfit the small database of synthesized catalysts.

2.3.3. Efficient Global Optimization and Design of Experiment

The regressors discussed in the previous section will in general result in a function potentially presenting various local maxima and minima that prevent efficient navigation of the search space purely based on optimization of the ML model. Therefore, we base our approach on global optimization methods [51] that use acquisition functions (selectors) that balance exploitation and exploration over the entire search space. The selectors are an important component of the DoE tool as they make optimal choices of which catalysts to investigate experimentally next and decrease the total number of measurements needed. Here, we combine the developed ML models discussed above with three different selectors algorithms, namely, Probability of Improvement (PI), Expected Improvement (EI), and Knowledgeable Gradient (KG). In addition, for evaluating the performance of each of these acquisition functions in our problem we also consider a random selector, which stochastically chooses samples without leveraging the statistical knowledge obtained via the ML models and the uncertainty quantification.

In the probability of improvement acquisition function, we check whether a so-far unexplored catalyst material is worth investigating experimentally by computing the probability that this new composition will provide an improvement over the currently known best-performing sample. Formally, the improvement function $I(\{x_i\})$ associated with a sample characterized by a set of input parameters $\{x_i\}$ is given as $I(\{x_i\}) = \max(ORR(\{x_i\}) - ORR(\{x_i^*\}), 0)$, where $ORR(\{x_i\})$ is an unknown 'black-box' function that describes the ground-truth correlations between the input

descriptors and the ORR activity, and $\{x_i^*\}$ represents the set of synthesis conditions indexing the best catalyst currently known. In general, we know the ground-truth function only over limited regions of the search space where measurements have already been performed, making calculations of the improvement function over uncharted systems impractical. Here, the machine learning models and uncertainty quantification approach play a decisive role in our design of experiment toolbox. We can evaluate the improvement function by assuming that $ORR(\{x_i\})$ can be surrogated by a normal distribution function with mean value $\mu(\{x_i\})$ and standard deviation $\sigma(\{x_i\})$, as obtained directly from the ML model in the case of the GPR algorithm, or computed via bootstrap statistical sampling for the cases of RF and ANN, as previously discussed. Under these assumptions, one can demonstrate that the probability of improvement is given by $PI(\{x_i\}) = \Phi(z)$, where $z = \frac{\mu(\{x_i\})-\mu(\{x_i^*\})-\epsilon}{\sigma(\{x_i\})}$, ϵ is an ad-hoc small positive number that controls the balance between exploitation and exploration, and $\Phi(z)$ represents the normal cumulative distribution function [52]. In our training database we observed that the activity of samples can be reproduced with a standard deviation of the order of ~ 1 mA/mg (see discussion in **Section 3**), and we decided to set ϵ to be one-tenth of this value to enable some exploration of the search space. Similarly, the expected improvement acquisition function is given by $EI(\{x_i\}) = \sigma(\{x_i\}) \times [\varphi(z) + z\Phi(z)]$ where $\varphi(z)$ is the normal probability density function [51]. Finally, the knowledge gradient function is a variation of the EI selector, obtained by considering $z = \frac{\mu(\{x_i\})-\mu(\{x_i^{**}\})-\epsilon}{\sigma(\{x_i\})}$, where $\{x_i^{**}\}$ is the set of synthesis conditions that maximize μ over the entire search space (i.e., explored and unexplored compositions) [53]. EI and KG are explicitly directly proportional to the uncertainty of predictions and automatically account for the trade-off between exploitation and exploration without the need for an ad-hoc parameter, thus we set $\epsilon = 0$ in these

cases. Regardless of the acquisition function of choice, trapping in any local maxima is prevented as the algorithm moves to areas of the search space with higher uncertainty at the cost of selecting catalysts with potential lower activity.

2.3.4. Regressor-Selector Combinations for Optimum Learning

Considering the three machine learning regressor models and three acquisition function selectors discussed above, there are a total of nine potential combinations that could be integrated to develop a design of experiment tool for selecting the next synthesis steps to test. However, there is no universal optimizer, as demonstrated in Ref. [54]. This implies that we do not know which regressor-selector pair will perform better with our starting datasets and will result in an optimized learning pathway. To choose the best ML model-acquisition function combination for our problem, we apply the design loop described in **Section 2.1**, constrained to the already synthesized and characterized electrocatalyst samples. Our approach is based on counting the average number of iterations needed for a given regressor-selector pair to find the best set of synthesis conditions $\{x_i^*\}$ within the initial dataset when starting from n_0 randomly selected samples in the training data. Specifically, our proposed methodology works as follows. From the original database with a total of N measured catalysts we stochastically choose $n_0<N$ samples. We count one iteration if the best material composition occurs in the initial random selection. When this does not happen, we train a chosen ML model on the n_0 samples and use a given acquisition function to pick the next catalyst from the remaining $N–n_0$ ones. The same regressor-selector pair is then used again, now starting from $n_0 +1$ samples and picking a new one from the $N – n_0 –1$ left over subset. The process continues until the highest performing catalyst $\{x_i^*\}$ is found and the total number of iterations n (including the initial random selection) is recorded. We repeat this approach 1,000 times starting with varying sets of n_0 randomly chosen catalysts and compute the average number of iterations

$\langle n \rangle$. The dependence of $\langle n \rangle$ versus n_0 for each ML model-acquisition function combination is then compared to that of the random selector for deciding the best optimization pair.

3. Results and Discussion

The initial mass activity for thirty-six catalyst samples obtained using specific synthesis and pyrolysis parameters is shown in **Tables S1** and **S2**. During measurements of the ORR activity of each catalyst, two cycles of the steady-state ORR polarization curves were recorded as a measure of the short-term stability. Multiple samples of the catalysts that showed the highest ORR activity have been synthesized and characterized to estimate the uncertainty in the ORR activity and to check the reproducibility of the experiments. It was observed that ORR measurements can be replicated with a standard deviation of about ~ 1 mA/mg, corresponding to a relative error in the range of 6-8% for the best performing catalysts. The ORR_1 and ORR_2 traces in **Fig. 3a,b** represent the averaged data of the forward and backward scans of the first cycle and second cycle, respectively, as an example for entry #5 for the furnace-open dataset. By comparing the variations of the ORR mass activity and/or H_2O_2 yields between these two cycles, the short-term stability can be estimated and evaluated for different catalysts. The well-defined peaks in the CV in **Fig. 3c** can be attributed to the Fe^{3+}/Fe^{2+} redox couple of the Fe-N-C catalysts [55, 56]. Integration of Fe redox peaks in the CV yields the Fe redox charge, a measure of the number of Fe sites, assumed to be the active sites or associated with the active sites in the Fe-N-C electrocatalyst. From the double-layer charging current in the potential range of 0.1 to 0.4 V vs. RHE, the capacitance can be calculated, which can be used to estimate the surface area of the catalyst that is accessible to the electrolyte [57-60]. The ORR catalytic activity of Fe-N-C catalysts is influenced by various physicochemical properties of the material, such as the concentration of $Fe\text{-}N_x$, the presence of less active species, such as Fe nanoparticles or Fe_3C, and the microporosity which is determined

by the extent of carbon graphitization [9, 17, 61-63]. All of the synthesis parameters studied here can influence these properties. For instance, only a limited number of ultra-fine FeO_x clusters formed during the heat treatment of Fe precursors can transform into atomically dispersed Fe-N_x, likely limited by the number of N-containing vacancies in the carbon structure formed during heat treatment of ZIF-8 [64]. Furthermore, excess Fe can react with the abundant carbon in the system to form carbide. Iron is also known to enhance graphitization, contributing to the formation of Fe_3C [65]. A higher pyrolysis temperature and longer pyrolysis times can increase the desorption of N from the N-C, thus leading to agglomeration of Fe into nanoparticles in conjunction with a loss of microporosity and surface area.

Fig. 4 shows the Pearson correlation coefficients using the initial datasets with eighteen samples for the furnace-closed cooling condition (**Fig. 4a**) and for the furnace open cooling condition (**Fig. 4b**). Here, the magnitude of the positive (negative) numbers indicates the degree of direct (inverse) correlation of the pairs. Parameters that can be controlled independently are uncorrelated and should result in a close to zero Pearson coefficient. This is indeed the case for many pairs of input material descriptors (e.g., Fe and phenanthroline-to-ZIF ratio (Phen); T_p and H_{rate}), although in some cases higher values have been observed for the Pearson coefficient (e.g., Fe and t_P for furnace closed; T_p and Phen for both furnace states). Such spurious correlations result from the limited amount of data available and are expected to approach zero as more samples covering wider regions of the search space are synthesized and characterized. Of particular interest are the relationships of the input parameters with the outcome activities ORR_1 and ORR_2, measured in the first and second RRDE potential cycles. Notice, however, that ORR_1 and ORR_2 appear to scale directly with each other in a near-perfect linear relation (additional discussion of this correlation is provided below). Hence, only one of these two properties is needed when

training ML models to establish correlations with the input descriptors. We will use ORR_1 from here on. The insets in **Fig. 4** highlight the corresponding correlation coefficients for ORR_1 with Fe precursor concentration (Fe), phenanthroline-to-ZIF ratio (Phen), pyrolysis temperature (T_P), dwell time at the pyrolysis temperature (t_P), and initial heating rate of the profile (H_{rate}). It should be noted that the ranges of values used in this parameter space were pre-selected to known regions of higher activity from the literature and so extrapolating beyond these ranges may significantly alter correlation values.

For both cooling cases, t_P is the most directly correlated parameter with increased activity. This speaks to the importance of accounting for the full pyrolysis (temperature versus time) curve beyond the typically reported controlled, single value pyrolysis temperature. The latter temperature is found to be either weakly directly or weakly inversely correlated with the initial catalytic activity. This may at least in part be due to previously reported nonlinear "peaked" behavior [66] where activity increases up to a given temperature and begins to decrease at higher temperatures. This contrasts with our previous findings [32], where pyrolysis temperature (without pyrolysis time considered) was found to be the dominant driver of ORR activity. The effect of the iron-precursor content effects is as reported previously, with activity improving with the Fe loading up to a certain point, with little or no activity increase occurring for higher loadings [32, 67]. Interestingly, the effect of the phenanthroline content on activity is directly correlated in the case of furnace closed and inversely correlated with furnace open. This suggests that ideal precursor compositions (at least in the case of the nitrogen precursor) couple to the pyrolysis curve, in particular the cooling rate, and that simultaneous optimization of precursor ratios and heat treatment approaches are needed to enhance ORR activities. In other words, iterative approaches to activity optimization with a single parameter being optimized at a time may miss key trends and

thus fail to meet the true potential of an electrocatalyst system. Finally, the importance of heating rate appears to be quite minimal in both cases but is more pronounced when rapid cooling/quenching of the system (furnace open) is avoided, thus suggesting that this is the least impactful parameter of the full pyrolysis curve.

Fig. 5 summarizes our findings regarding the performance of the various regressor-selector combinations for the two cooling states of the furnace used during the synthesis process. **Fig. 5a** shows a schematic representation of the methodology described in section 2.3.4, in which we confined our adaptive learning loop to the subspace of initially measured catalysts and counted the average number of iterations $\langle n \rangle$ that are necessary for a given machine learning-acquisition function pair to find the highest ORR_1 activity within this dataset when starting from n_0 randomly chosen samples. In **Fig. 5a,b** the blue circle markers show the numerically calculated values of $\langle n \rangle$ for the baseline random selector, which always chooses the next sample stochastically, regardless of any statistical information acquired by the machine learning model. In this case, the average number of iterations can be derived analytically as $\langle n_{random} \rangle = 1 + (N - n_0) \times (N - n_0 + 1)/2N$ (dashed blue line in the plots), which agrees well with computational simulations. Here, N (= 18 in our system) is the total number of samples in the entire training database for each cooling rate scenario. Note that because we always count the very first choice of samples as one iteration in our algorithm, then $\langle n_{random} \rangle = 1$ only when $n_0 = N$.

We expect that if n_0 is large enough to cover most of the search space, all regressor-selector combinations should perform similarly to the random selector. Indeed, in the limit when the initial number of training samples is $n_0 = N - 1$, the chances of not finding the best catalyst in the initial selection is $1/N$ and all machine learning model-acquisition function pairs have 100% probability of discovering the optimum sample in the next iteration. In the opposite limit (when n_0 is small),

the performance of all regressor-selector pairs deteriorates as the machine learning models may fail to properly capture the correlations between input descriptors and outcome properties. Intermediate values of n_0 are therefore the most appropriate for deciding the best regressor-selector pair for our systems, and for this reason, we limit the number initial of random picks to the range shown in the figures ($3 \leq n_0 \leq 10$). As presented in **Fig. 5a,b**, $\langle n \rangle$ generally decreases with increasing n_0 as the probability of finding the best sample in the first random selection is n_0/N. Consistent with the inexistence of a universal optimizer [54], we observe that there are algorithms that underperform the random selector. For the samples synthesized under slow cooling conditions (furnace closed), **Fig. 5b** shows that all regressor-selector combinations involving the developed ANN model (diamond markers) should be avoided. We also note that GPR-PI and GPR-EI perform worse than stochastic selections when $n_0 < 5$, but their efficiency becomes more promising as n_0 increases. All other possibilities of machine learning models-acquisition functions outperform the random selector, and we choose to work with the random forest-expected improvement design of experiment tool for the furnace-closed case, since it consistently finds the best sample in the training dataset with fewer average iterations than any other pair in the entire range of values considered for n_0. Strikingly, **Fig. 5c** reveals that the RF-EI combination is the least promising approach for navigating the search space in the case of fast cooling rates (furnace open). In this situation, we observe that any search tool involving the KG global optimization method (red curves) is inappropriate, as their performance deteriorates when more initial samples n_0 are considered. Finally, it should be noted that although RF-PI appears to be the best option in **Fig. 5c** at small values of n_0, it is slightly outperformed by ANN-EI and ANN-PI when larger sets of initial data ($n_0 > 7$) are considered. As there are no reasons for not using the complete dataset of available samples for training ML models in our problem, we have selected ANN-PI (which is always better

than ANN-EI in the range considered) as the DoE tool to be used in the case of high-throughput synthesis with the furnace open. We comment that the shift in the ideal approach depending on the furnace state further confirms the inexistence of a universal optimizer in a practical system [54], and highlights that the use of different cooling rates during the synthesis process strongly affects the landscape of the parameter's search space.

In **Fig. 6,** we discuss the performance of the selected machine learning algorithms trained on our existing databases and how the computational models can be leveraged to investigate the effect on the catalyst activity of the synthesis conditions beyond the linear regime considered by the Pearson correlation coefficients in **Fig. 4**. Following the discussion in the previous paragraph, we will concentrate on the random forest and artificial neural network regressor. **Fig. 6a,b** compare the ORR_1 predictions, obtained with the hyperparameters-optimized RF and ANN models (see **Section 2.3.2**), for the cases of furnace closed and open, respectively, with the measured values in the initial dataset. The models accurately capture correlations between the independently controlled parameters and the ORR activity, although they tend to slightly underestimate the experimental values of ORR_1 for the best-performing samples. Also, there is considerable uncertainty in the model predictions due to the sparse availability of data, resulting in standard deviations as high as ~2 mA/mg (*ca.* 20% relative error) for some samples. This indicates that further exploration of the search space is still needed, and the design of experiment framework may suggest new materials that would decrease the overall uncertainty of the models before exploiting around local maxima of the activity. Moreover, it is worth noticing that overfitting does not occur due to the statistical bootstrap sampling strategy employed for modeling predictions and uncertainty quantification.

In **Fig. 6c,d** we quantify and compare the dependence of ORR_1 on five material descriptors using nonlinear correlations between input-output features, as captured by the RF and ANN models for the slow and fast cooling scenarios, respectively. Our sensitivity analysis is based on the permutation importance score, which works as follows. We create synthetic datasets from the training database by randomly shuffling the values of a single input parameter, thus breaking the relationship between that specific feature and the outcome property of interest. We then evaluate the decrease in an ML model metric relative to the original (unshuffled) database. The larger the change achieved in the metric, the more the model depends on the feature under consideration. Clearly, the sensitivity of the ORR activity on the synthesis descriptors depends on the metric used to compute the permutation importance score. In our calculations, we perform the above procedure over 1,000 random shuffles for each synthesis parameter and we report the normalized average and standard deviation for the RMSE, coefficient of determination (R^2), and mean-absolute-percentage error (MAPE) metrics. Note that the computed average MAPE values are parameter-dependent, but their large uncertainty prevents us from effectively using it to distinguish the relative relevance of the synthesis conditions. A different scenario occurs for the other two metrics, where we can distinctly separate the dominant synthesis descriptors. Similar to the Pearson correlation coefficient calculation, we notice that RMSE and R^2 both rank the pyrolysis time as the most important parameter affecting the catalyst activity, regardless of the cooling process. For the furnace closed case, these two metrics also suggest that ORR_1 is equally sensitive to the Fe loading, heating rate, and pyrolysis temperature, despite small variations in their permutation importance score. Moreover, unlike the observations from the Pearson coefficient, the phenanthroline-to-ZIF ratio seems to play a minor role in this case, highlighting the nonlinear nature of the correlations in our system. For the furnace open case, all parameters (except for t_P) similarly affect ORR_1, with

the average values of the metrics implying that the sensitivity ranking of the descriptors is a phenanthroline-to-ZIF ratio, Fe loading, heating rate, and pyrolysis temperature. We mention that as future experimental data is integrated in the training database one expects a significant decrease in the uncertainty of the permutation importance score, allowing for a more robust sensitivity analysis regardless of the metric of choice. Altogether, these results emphasize that the influence of synthesis conditions on ORR_1 is highly nontrivial, making the use of Edisonian trial-and-error approaches unlikely to succeed in a timely fashion for the discovery of optimal synthesis conditions of Fe-N-C electrocatalysts.

Fig. 7 shows the measured ORR mass activities for all samples synthesized in this work. For easy presentation, we have arbitrarily labeled the samples in the initial datasets with a numeric index from 1-18, corresponding to the order they are presented in the supporting material (**Tables S1** and **S2**). We note that in the initial database the two best-performing samples had activities of 12.3 mA/mg and 10.8 mA/mg for the furnace closed and open, respectively. These activities are achieved for an Fe loading of 1.60%, phenanthroline-to-ZIF ratio of 0.00%, pyrolysis time of 2.0 h, and heating rate of 5.0 ºC/min in both cases, while the pyrolysis temperature was set to 1050 ºC in the former and 950 ºC in the latter case. A total of fourteen new catalysts (seven for each of the two cooling rates considered) were synthesized as we iterated through the adaptive learning framework loop and used the developed design of experiment tools to make optimal selections of synthesis conditions to test next. The measured ORR mass activity for these samples, as well as the ML model predictions and their uncertainties, are shown in **Fig. 7** with labels in the range of 19-25 (see synthesis conditions in the supporting information, **Table S3** and **S4**), corresponding to the order that the samples have been prepared and characterized. The plot shows that the two trained ML models (RF and ANN) are effectively able to predict the activities of unexplored

compositions regardless of the cooling process used for catalyst preparation. Indeed, the computational calculations agree, within the standard deviations, with almost all experimentally measured ORR_1 values (the only exception being sample #22 for the furnace-open scenario). For some catalysts, the statistical uncertainty associated with the ML predictions is substantially high, which is a direct consequence of the limited number of samples in the initial dataset. Future iterations through the adaptive learning loop should provide ORR activity values from poorly explored regions of the search space, resulting in more accurate ML models with reduced uncertainty.

Note that the second sample suggested by the adaptive learning framework (sample #20) for the furnace-open case resulted in $ORR_1 = 16.3 \pm 0.4$ mA/mg that corresponds to an activity enhancement of about 51% with respect to the best-performing catalyst prepared under the same cooling conditions. This value is also 33% higher than the previously observed global maximum, which occurred for the furnace closed. We comment that this activity is about seven times that of state-of-the-art pyrolyzed Fe-N-C catalysts reported in other works (e.g., $ORR_1 = 2.25 \pm 1.25$ mA/mg, measured at 0.8 V via RRDE using a catalyst loading of 0.8 mg/cm^2, have been reported in [68]). Interestingly, the newly discovered best sample has been prepared with a significantly lower Fe loading in the precursors (0.80%) and a slightly higher pyrolysis time (t_P = 3.0 h) than the two previously measured materials with the largest ORR_1. The other synthesis parameters are phenanthroline-to-ZIF ratio = 0.00%, T_p = 1050 ºC, and H_{rate} = 5.0 ºC/min. The observed striking increase in the ORR activity demonstrates how our advanced theory-experiment framework can accelerate the discovery of electrochemical materials with enhanced functionalities and benefit fuel cell and catalysis research. The optimum Fe loading and pyrolysis temperature for this sample are comparable to those reported in Ref. [17]. They found an optimized Fe loading of 1 wt.%

(amongst 0.5, 1 and 1.5 wt.% Fe) and T_p = 1050 ºC. On the other hand, our best performing catalyst has been synthesized while opening the furnace at the end of pyrolysis. This is consonant with a few studies that have reported beneficial effects by quenching the temperature after pyrolysis as a way to decrease desorption of doped N atoms, but also formation of Fe-based nanoparticles [69, 70], thus decreasing Fe–N–C catalytic site density [71].

Despite the good predictive capabilities of the developed regression models, we notice substantial differences in the performance of the adaptive learning framework for optimizing Fe-N-C -based catalysts subjected to slow or fast cooling rates. In fact, in the former case, no new sample has been discovered with higher activity than the best ones in the original databases. In the latter case, four out of the seven catalysts prepared using the computational model guidance demonstrated an improvement in ORR activity. There are a few reasons that could explain this asymmetry in the two optimization processes. The simplest is that when randomly drawing the eighteen samples that compose the initial database for the furnace closed situation, we have already and accidentally discovered quasi-optimal synthesis conditions, in which case any new sample will likely underperform the best known so far. However, this is an unlikely explanation given the size of the initial dataset and the total number of possible synthesis steps available. A second plausible explanation is that the synthesis conditions for the training datasets for the furnace closed and open are intrinsically different (see data in the Supporting Information), and there are no compelling reasons why the chosen best regressor-selector pair optimizers should be equally efficient in both cases. An inspection of **Fig. 5b,c** reveals that the ANN-PI combination in the case of the furnace open slightly outperforms the RF-EI pair in the case of the furnace closed for finding the best catalyst in their corresponding databases. Finally, the use of distinct acquisition functions in each case, although well justified by our methodology, is certain to influence how the optimization

algorithms navigate the high-dimensional search space of uncharted synthesis conditions. Indeed, in the expected improvement function, the ML model uncertainty directly controls the exploitation-exploration trade-off. Due to the sparse data distribution, the uncertainty of model predictions is typically larger in the search space of unexplored combinations than in the original training set. As a result, the RF-EI design of experiment tool (furnace closed) may iteratively suggest testing new systems with lower activity but higher uncertainty to improve the overall performance of the ML model over the entire search space in the long term. On the other hand, in the probability of improvement acquisition function, an arbitrarily set parameter ϵ assesses the merits of exploitation versus exploration. Our choice for the value of ϵ (one-tenth of typical ORR_1 standard deviation observed experimentally, see **Section 2.3.3**) does not seem to significantly influence the optimization approach since RF-EI (furnace closed) and ANN-PI (furnace open) show similar performance when applied to their respective training data (see **Fig. 5b,c**). However, this fixed value of ϵ is smaller than the uncertainties observed in the uncharted search space, and this may have favored initial exploitation around a known local maximum for the furnace open case during the first iterations through the learning loop. In any scenario, further exercise through our adaptive design framework is likely needed to fully elucidate whether new catalysts with higher activity, synthesized under slow cooling conditions, are possible.

In **Fig. 8** we plot the ORR activity (ORR_2) measured in the second RRDE cyclic voltammogram scan versus that obtained in the ORR measurement (ORR_1). The former scales linearly with the latter regardless of the cooling rate employed during the synthesis process, as already captured by the Pearson correlation coefficient matrix in **Fig. 4a,b**. This result strongly suggests that the short-term stability (ORR activity loss during the first voltametric measurement) of Fe-N-C -based catalysts is largely unaffected by the specific synthesis conditions. We formally

capture the correlations between ORR_2 and ORR_1 by training an ML linear regression algorithm using all thirty-six samples initially synthesized and by using statistical bootstrap sampling to quantify the uncertainty. The resulting model is $ORR_2 = (0.91 \pm 0.03) \times ORR_1 - (0.32 \pm 0.19)$, which has a coefficient of determination $R^2 = 0.95$ when we consider the fourteen samples obtained via adaptive learning as the testing database. As depicted in **Fig. 8a**, this model effectively describes the relationship between ORR_2 and ORR_1 for all samples available within a 99% confidence interval. It is interesting to note that the intercept of the linear regression model is non-zero ($= 0.32 \pm 0.19$ mA/mg). This implies that for poorly performing catalysts we need to consider nonlinear terms to accurately describe the short-term stability since ORR_2 should vanish in the (unrealistic) limit of $ORR_1 \rightarrow 0$. This regime, however, is the least relevant as we are typically looking for optimum materials and conditions that maximize the catalyst activity for practical applications. In the opposite limit, one can neglect the intercept term in the linear model and approximate the short-term stability as $Stability\,(\%) = 100 \times (ORR_2/ORR_1 - 1) \approx (9 \pm 3)\%$. Data in **Fig. 8b** confirm the model predictions that samples with ORR_1 larger than ~8 A/g will lose about 10% of their initial activity from the first to second RRDE cyclic voltammetric scans.

4. Conclusions

In summary, we have successfully developed and demonstrated an adaptive learning framework for accelerated design and discovery of Fe-N-C oxygen reduction reaction electrochemical catalysts with enhanced ORR activity. Our approach leverages high-throughput synthesis using commercial precursors, machine learning, uncertainty quantification, and global optimization algorithms to realize a design-of-experiment tool that makes optimum selections of high-throughput synthesis conditions and limited the number of often expensive and time-

consuming experiments. We have used the developed machine learning regression models to determine the critical roles played by various independently controlled synthesis parameters on ORR activity and short-term stability, including iron loading, phenanthroline-to-ZIF-8 ratio, pyrolysis temperature, pyrolysis time, and heating and cooling rates. By employing the adaptive design methodology with an initial dataset of only thirty-six samples prepared under unique conditions, over a universe of nearly 10^{10} possibilities in a six-dimensional search space, we have guided the synthesis of fourteen unexplored materials. We demonstrated that four of the new samples showed higher ORR activity, with one of them outperforming by more than 33% the previously known best catalyst in the entire training database. The sample with the highest ORR activity of 16.3 mA/mg (0.8 wt.% Fe, phenanthroline-to-ZIF ratio = 0.00%, T_p = 1050 ºC, t_P = 3.0 h, H_{rate} = 5.0 ºC/min, open furnace) showed a 600% activity improvement relative to the highest activity values reported in the literature for the same class of materials [68]. Furthermore, we have demonstrated the importance of the pyrolysis temperature profile (not just the pyrolysis temperature hold) and how it can couple to other parameters to influence the resulting ORR activity. Despite the overall success of the adaptive learning strategy, we mention that interpreting the computational DoE tool results and recommending samples to test next is cumbersome. The use of physics/chemistry-informed ML approaches [72] may potentially help in circumventing this challenge while providing further improvements in the performance of existing models. While enhancing adaptive learning capabilities with physics-informed ML for catalyst synthesis optimization could lead to promising research directions, physical or chemical models elucidating how variations in the synthesis process affect the activity of pyrolyzed Fe-N-C catalysts are still in need of development. Finally, our adaptive learning approach can be readily translated to benefit

research involving other materials and systems of interest to clean energy sciences, beyond the catalytic systems considered here for the oxygen reduction reaction.

Acknowledgements

The work was funded by the U.S. Department of Energy (DOE), Energy Efficiency and Renewable Energy, Hydrogen and Fuel Cell Technologies Office under the auspices of the Electrocatalysis Consortium (ElectroCat). The authors would like to thank the ElectroCat DOE program managers, Dimitrios Papageorgopoulos, David Peterson, and McKenzie Hubert. We thank Yuri Muniz for valuable discussions on this project's various data science and machine learning aspects. This work was authored in part by Los Alamos National Laboratory operated by Triad National Security, LLC under US DOE contract no. 89233218CNA000001, by Argonne National Laboratory, managed by the University of Chicago Argonne, LLC under contract DE-AC-02-06CH11357.

Glossary of Acronyms and Symbols

ANN – artificial neural network

CV – cyclic voltammogram

DoE – design of experiment

EI – expected improvement

$E_{½}$– half-wave potential

GC – glassy carbon

GPR – gaussian process regressor

H_{rate} – heating rate

i_{lim} – limiting current

KG – knowledgeable gradient

MAPE – mean-absolute-percentage error

ML – machine learning

ORR – oxygen reduction reaction

PEFC – polymer electrolyte fuel cell

PGM – platinum group metal

Phen – phenanthroline-to-ZIF-8 ratio

PI – probability of improvement

PTFE – polytetrafluoroethylene

RF – random forest

RHE – reversible hydrogen electrode

RMSE – root-mean-square error

RRDE – rotating ring disk electrode

R^2 – coefficient of determination

t_P – pyrolysis time

T_p – pyrolysis temperature

ZIF – zeolitic imidazolate framework

References

1. I. Staffell, D. Scamman, A. V. Velazquez Abad, P. Balcombe, P. E. Dodds, P. Ekins, N. Shahd, K. R. Ward, *The role of hydrogen and fuel cells in the global energy system*. Energy Environ. Sci. **12** (2019) 463-491.

2. D. A. Cullen, K. C. Neyerlin, R. K. Ahluwalia, R. Mukundan, K. L. More, R. L. Borup, A. Z. Weber, D. J. Myers, A. Kusoglu*, New roads and challenges for fuel cells in heavy-duty transportation*, Nat. Energy **6** (2021) 462-474.

3. H. Yang, X. Han, A. I. Douka, L. Huang, L. Gong, C. Xia, H. S. Park, and B. Y. Xia, *Advanced oxygen electrocatalysts in energy conversion and storage*. Adv. Func. Mater. **31** (2021) 2007602.

4. B. Pivovar, *Catalysts for fuel cell transportation and hydrogen related uses*. Nat. Catal. **2**, (2019) 562−565.

5. S. Thompson, D. Papageorgopoulos, *Platinum group metal-free catalysts boost cost competitiveness of fuel cell vehicles*. Nat. Catal. **2** (2019) 558−561.

6. D. A. Cullen, K. C. Neyerlin, R. K. Ahluwalia, R. Mukundan, K. L. More, R. L. Borup, A. Z. Weber, D. J. Myers, A. Kusoglu, *New roads and challenges for fuel cells in heavy-duty transportation*. Nat. Energy **6** (2021) 462−474.

7. F. T. Wagner, B. Lakshmanan, M. F. Mathias, *Electrochemistry and the future of the automobile*. J. Phys. Chem. Lett. **1** (2010) 2204−2219.

8. M. Lefèvre, E. Proietti, F. Jaouen, J.-P. Dodelet, *Iron-based catalysts with improved oxygen reduction activity in polymer electrolyte fuel cells*. Science **324** (2009) 71−74.

9. A. Zitolo, V. Goellner, A. Armel, M.-T. Sougrati, T. Mineva, L. Stievano, E. Fonda, F. Jaouen, *Identification of catalytic sites for oxygen reduction in iron- and nitrogen-doped graphene materials*. Nat. Mater. **14** (2015) 937−944.

10. M. Shao, Q. Chang, J.-P. Dodelet, R. Chenitz, *Recent advances in electrocatalysts for oxygen reduction reaction*. Chem. Rev. **116** (2016) 3594−3657.

11. H. Zhang, S. Hwang, M. Wang, Z. Feng, S. Karakalos, L. Luo, Z. Qiao, X. Xie, C. Wang, D. Su, Y. Shao, G. Wu, *Single atomic iron catalysts for oxygen reduction in acidic media: particle size control and thermal activation*. J. Am. Chem. Soc. **139** (2017) 14143−14149.

12. Y. Chen, S. Ji, C. Chen, Q. Peng, D. Wang, Y. Li, *Single atom catalysts: synthetic strategies and electrochemical applications*. Joule **2** (2018) 1242−1264.

13. F. Jaouen, D. Jones, N. Coutard, V. Artero, P. Strasser, A. Kucernak, *Toward platinum group metal-free catalysts for hydrogen/air proton-exchange membrane fuel cells*. Johnson Matthey Technol. Rev. **62** (2018) 231−255.

14. X. Wan, X. Liu, Y. Li, R. Yu, L. Zheng, W. Yan, H. Wang, M. Xu, J. Shui, *Fe−N−C electrocatalyst with dense active sites and efficient mass transport for high-performance proton exchange membrane fuel cells*. Nat. Catal. **2** (2019) 259−268.

15. C. Wan, X. Duan, Y. Huang, *Molecular design of single-atom catalysts for oxygen reduction reaction*. Adv. Energy Mater. **10** (2020) 1903815.

16. L. Osmieri, J. Park, D. A. Cullen, P. Zelenay, D. J. Myers, K. C. Neyerlin, *Status and challenges for the application of platinum group metal-free catalysts in proton exchange membrane fuel cells*. Current Opinion Electrochem. **25** (2021) 100627−100638.

17. E. Proietti, F. Jaouen, M. Lefèvre, N. Larouche, J. Tian, J. Herranz, J.-P. Dodelet, *Iron-based cathode catalyst with enhanced power density in polymer electrolyte membrane fuel cells*. Nat. Comm. **2** (2011) 416.

18. H. Zhang, H. Osgood, X. Xie, Y. Shao, and G. Wu, *Engineering nanostructures of PGM-free oxygen-reduction catalysts using metal-organic frameworks*. Nano Energy **31** (2017) 331−350.

19. X. Tian, X. F. Lu, B. Y. Xia, X. W. Lou, *Advanced electrocatalysts for the oxygen reduction reaction in energy conversion technologies*. Joule **4** (2020) 45–68.

20. S. Liu, C. Li, M. J. Zachman, Y. Zeng, H. Yu, B. Li, M. Wang, J. Braaten, J. Liu, H. M. Meyer III, M. Lucero, A. J. Kropf, E. E. Alp, Q. Gong, Q. Shi, Z. Feng, G. Wang, H. Xu, D. J. Myers, J. Xie, D. A. Cullen, S. Litster, G. Wu, *Atomically dispersed iron sites with a nitrogen-carbon coating as highly active and durable oxygen reduction catalysts for fuel cells*. Nat. Energy **7** (2022) 652–663.

21. V. Armel, S. Hindocha, F. Salles, S. Bennett, D. Jones, *Structural descriptors of zeolitic−imidazolate frameworks are keys to the activity of Fe−N−C catalysts*. J. Am. Chem. Soc. **139** (2017) 453–464.

22. M. Chen, Y. He, J.S. Spendelow, G. Wu, *Atomically dispersed metal catalysts for oxygen reduction*. ACS Energy Lett. **4** (2019) 1619–1633.

23. M. Lefèvre, J.-P. Dodelet, P. Bertrand, *Molecular oxygen reduction in PEM fuel cells: evidence for the simultaneous presence of two active sites in Fe-based catalysts*. J. Phys. Chem. B **106** (2002) 8705–8713.

24. M. K. Jeon, C. H. Lee, G. Il Park, K. H. Kang, *Combinatorial search for oxygen reduction reaction electrocatalysts: A review*. J. Power Sources **216** (2012) 400–408.

25. K. T., Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Machine learning for molecular and materials science*. Nature **559** (2018) 547−555.

26. L. Himanen, A. Geurts, A. S. Foster, P. Rinke, *Data-Driven Materials Science: Status, Challenges, and Perspectives*. Adv. Sci. **6** (2019) 1900808.

27. D. P. Tabor, L. M. Roch, S. K. Saikin, C. Kreisbeck, D. Sheberla, J. H. Montoya, S. Dwaraknath, M. Aykol, C. Ortiz, H. Tribukait et al., *Accelerating the discovery of materials for clean energy in the era of smart automation.* Nat. Rev. Mater. **3** (2018) 5−20.

28. G. H. Gu, J. Noh, I. Kim, Y. Jung, *Machine learning for renewable energy materials*. J. Mater. Chem. A **7** (2019) 17096−17117.

29. C. Chen, Y. X. Zuo, W. K. Ye, X. G. Li, Z. Deng, S. P. Ong, *A Critical Review of Machine Learning of Energy Materials*. Adv. Energy Mater. **10** (2020) 1903242.

30. H. Mai, T. C. Le, D. Chen, D. A. Winkler, R. A. Caruso, *Machine learning for electrocatalyst and photocatalyst design and discovery*. Chem. Rev. **122** (2022) 13478–13515.

31. T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K. Shimizu, *Machine Learning for Catalysis Informatics: Recent Applications and Prospects.* ACS Catal. **10** (2020) 2260−2297.

32. M. R. Karim, M. Ferrandon, S. Medina, E. Sture, N. Kariuki, D. J. Myers, E. F. Holby, P. Zelenay, T. Ahmed, *Coupling High-Throughput Experiments and Regression Algorithms to Optimize PGM-Free ORR Electrocatalyst Systems*. ACS. Appl. Energy Mater. **3** (2020), 9083-9088.

33. R. Gomez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D. G. Ha, T. Wu et al., *Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach.* Nat. Mater. **15** (2016) 1120−1127.

34. P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y.-H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang, P. K. Herring, M. Aykol, S. J. Harris, R. D. Braatz, S. Ermon, W. C. Chueh, *Closed-Loop Optimization of Fast-Charging Protocols for Batteries with Machine Learning.* Nature **578** (2020) 397−402.

35. Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming, A. A. Lee, *Identifying Degradation Patterns of Lithium Ion Batteries from Impedance Spectroscopy Using Machine Learning*. Nat. Commun. **11** (2020) 1706.

36. H. Xu, J. Ma, P. Tan, B. Chen, Z. Wu,Y. Zhang, H. Wang, J. Xuan, M. Ni, *Towards online optimisation of solid oxide fuel cell performance: Combining deep learning with multi-physics simulation.* Energy Artif. Intell. **1** (2020) 100003.

37. R. Ding, Y. Ding, H. Zhang, R. Wang, Z. Xu, Y. Liu, W. Yin, J. Wang, J. Li, J. Liu, *Applying Machine learning to Boost the Development of High-Performance Membrane*

Electrode Assembly for Proton Exchange Membrane Fuel Cells. J. Mater. Chem. A **9** **(**2021) 6841−6850.

38. Z. Ghahramani, *Probabilistic machine learning and artificial intelligence*. Nature **521** (2015) 452–459.

39. D. Xue, P. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, *Accelerated search for materials with targeted properties by adaptive design*. Nat. Commun. **7** (2016) 11241.

40. M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C.-T. Dinh, P. De Luna, Z. Yu, A. S. Rasouli, P. Brodersen, S. Sun, O. Voznyy, C.-S. Tan, M. Askerka, F. Che, M. Liu, A. Seifitokaldani, Y. Pang, S.-C. Lo, A. Ip, Z. Ulissi, E. H. Sargent, *Accelerated Discovery of CO_2 Electrocatalysts Using Active Machine Learning.* Nature **581** (2020) 178−183.

41. T. Xing, J. Sunarso, W. Yang, Y. Yin, A. M. Glushenkov, L. Hua Li, P. C. Howlettb, Y. Chen, *Ball milling: a green mechanochemical approach for synthesis of nitrogen doped carbon nanoparticles*. Nanoscale **5** (2013) 7970–7976.

42. A. Shen, Y. Zou, Q. Wang, R. A. W. Dryfe, X. Huang, S. Dou, L. Dai, S. Wang, *Oxygen reduction reaction in a droplet on graphite: direct evidence that the edge is more active than the basal plane*. Angew. Chem. Int. Ed. **53** (2014) 10804–10808.

43. X. Chen, L. Yu, S. Wang, D. Deng, X. Bao, *Highly active and stable single iron site confined in graphene nanosheets for oxygen reduction reaction*. Nanoenergy **32** (2017) 353–358.

44. https://www.anl.gov/cse/highthroughput-research-laboratory

45. F. Jaouen, V. Goellner, M. Lefevre, J. Herranz, E. Proietti, J. P. Dodelet, *Oxygen reduction activities compared in rotating-disk electrode and proton exchange membrane fuel cells for highly active Fe-N-C catalysts.* Electrochim. Acta **12** (2013) 619-628.

46. G. Chon, M. Suk, F. Jaouen, M. W. Chung, C. H. Choi, *Deactivation of Fe-N-C catalysts during catalyst ink preparation process*. Cat. Today **359** (2021) 9-15.

47. U. A. Paulus, T. J. Schmidt, H. A. Gasteiger, R. J. Behm, *Oxygen reduction on a high-surface area Pt:Vulcan carbon catalyst: a thin-film rotating ring-disk electrode study*. J. Electroanal. Chem. **495** (2001) 134–145.

48. L. Osmieri, R. K. Ahluwalia, X. Wang, H. T. Chung, X. Yin, A. J. Kropf, J. Park, D. A. Cullen, K. L. More, P. Zelenay, D. J. Myers, K. C. Neyerlin, *Elucidation of Fe-N-C electrocatalyst active site functionality via in-situ X-ray absorption and operando determination of oxygen reduction reaction kinetics in a PEFC*. Appl. Catal. B: Environ. **257** (2019) 117929–117941.

49. https://scikit-learn.org/stable/

50. T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. 2nd. ed. Springer (New York), 2009.

51. D. R. Jones, M. Schonlau, W. J. Welch, *Efficient global optimization of expensive black-box functions*. J. Global Optim. **13** (1998) 455–492.

52. H. J. Kushner, *A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise*. Journal of Basic Engineering **86** (1964) 97-106.

53. W. B. Powell, I. O. Ryzhov, in Wiley Series in *Probability and Statistics,* Wiley (2013).

54. D. H. Wolpert, W. G. Macready, *No free lunch theorems for optimization*. IEEE Trans. Evol. Comput. **1** (1997) 67–82.

55. M. Elvington, H. T. Chung, L. Lin, X. Yin, P. Ganesan, P. Zelenay, H. R. Colón-Mercado, *Communication - On the lack of correlation between the voltammetric redox couple and ORR activity of Fe-N-C catalysts.* J. Electrochem. Soc. **167** (2020) 134510.

56. P. K. Sow, Z. Lu, H. Talebian, L. Damron, W. Merida, *Double layer capacitance measurements to characterize the water instruction into porous materials*. J. Phys. Chem. C **120** (2016) 24794–24802.

57. V. V. Panic, R. M. Stevanovic, V. M. Jovanovic, A. B. Dekanski, *Electrochemical and capacitive properties of thin-layer carbon black electrodes*. J. Power Sources **181** (2008) 186–192.

58. J-J. Jow, L.-Y. Hsieh, H.-P Cho, H.-R. Chen C.-W. Kuo, *Determination of surface area of carbon-black by simple cyclic voltammetry measurements in aqueous H_2SO_4*. J. Ind. Eng. Chem. **19** (2013) 1730–1734.

59. F. Beck, M. Dolata, E. Grivei, N. Probst, *Electrochemical supercapacitors based on industrial carbon blacks in aqueous H_2SO_4*. J. Appl. Electrochem. **31** (2001) 845–853.

60. T. A. Centeno, M. Hahn, J. A. Fernández, R. Kötz, F. Stoeckli, *Correlation between capacitances of porous carbons in acidic and aprotic EDLC electrolytes*. Electrochem. Comm. **9** (2007) 1242–1246.

61. Y. He and G. Wu, *PGM-free Oxygen-Reduction Catalyst Development for Proton-Exchange Membrane Fuel Cells: Challenges, Solutions, Promises*. Acc. Mater. Res. **3** (2022) 224-236.

62. X. Wang, H. Zhang, H. Lin, S. Gupta, C. Wang, Z. Tao, H. Fu, T. Wang, J. Zheng, G. Wu, X. Li, *Directly converting Fe-doped metal–organic frame works into highly active and stable Fe-N-C catalysts for oxygen reduction in acid*. Nano Energy **25** (2016) 110.

63. H. Zhang, S. Hwang, M. Wang, Z. Feng, S. Karakalos, L. Luo, Z. Qiao, X. Xie, C. Wang, D. Su, Y. Shao, G. Wu, *Single atomic iron catalysts for oxygen reduction in acidic media: particle size control and thermal activation.* J. Am. Chem. Soc. **139** (2017) 14143.

64. J. Li, H. Zhang, W. Samarakoon, W. Shan, D.A. Cullen, S. Karakalos, M. Chen, D. Gu, K.L. More, G. Wang, Z. Feng, Z. Wang, G. Wu, *Thermally driven structure and performance evolution of atomically dispersed FeN_4 sites for oxygen reduction.* Angew. Chem. Int. Ed. **58** (2019) 18971.

65. H. Marsh, A.P. Warburton, *Catalysis of Graphitization.* J. Appl. Chem. **20** (1970) 133.

66. G. Wu, K. L. More, C. M. Johnston, P. Zelenay, *High-performance electrocatalysts for oxygen reduction derived from polyaniline, iron and cobalt*. Science **332** (2011) 443-447.

67. H. Zhang, H. T. Chung, D. A. Cullen, S. Wagner, U. I. Kramm, K. L. More, P. Zelenay, G. Wu, *Atomically Dispersed Iron Catalysts for Polymer Electrolyte Fuel Cells*. Energy Environ. Sci. **12** (2019) 2548-2558.

68. M. Primbs, Y. Sun, A. Roy, D. Malko, A. Mehmood, M.-T. Sougrati, P.-Y. Blanchard, G. Granozzi, T. Kosmala, G. Daniel, P. Atanassov, J. Sharman, C. Durante, A. Kucernak, D. Jones, F. Jaouen, P. Strasser, *Establishing reactivity descriptors for platinum group metal (PGM)-free Fe-N-C catalysts for PEM fuel cells*. Energy Environ. Sci. **13** (2020) 2480.

69. A. Morozan, M. T. Sougrati, V. Goellener, D. Jones, L. Stievano, F. Jaouen, *Effect of Furfuryl alcohol on metal organic framework-based Fe/N/C electrocatalysts for polymer electrolyte membrane fuel cells.* Electrochimica Acta **119** (2014) 192–205.

70. S. Yasuda, Y. Uchibori, M. Wakeshima, Y. Hinatsu, H. Ogawa, M. Yano, H. Asaoka, *Enhancement of Fe-N-C carbon catalyst activity for the oxygen reduction reaction: effective increment of active sites by short and repeated heating process.* RSC Adv. **8** (2018) 37600-37605.

71. F. Jaouen, M. Lefevre, J.-P. Dodelet, M. Cai, *Heat-treated Fe/N/C catalysts for O_2 electroreduction: are active sites hosted in micropores?*. Phys. Chem. B **110** (2006) 5553–5558.

72. G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, *Physics informed machine learning*. Nat. Rev. Phys. **3** (2021) 422.

Figure 1

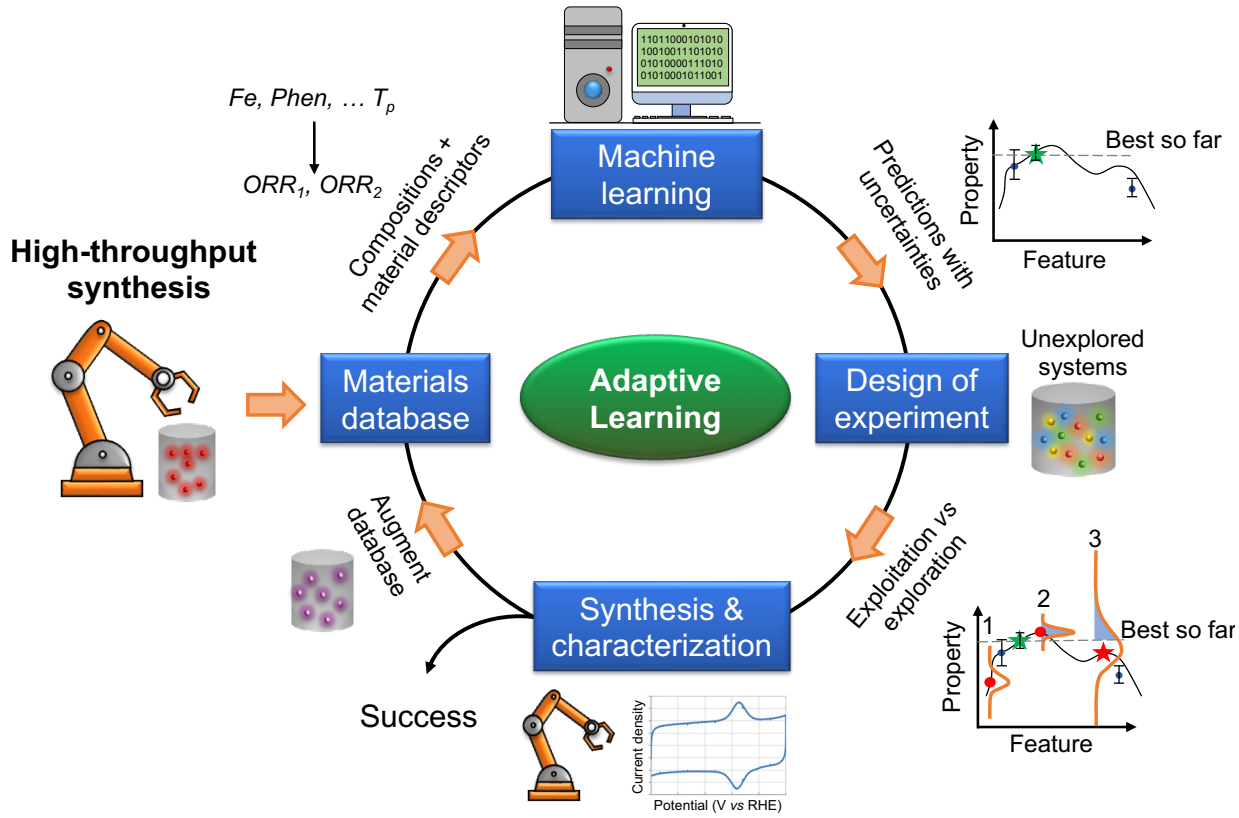

Fig. 1. Adaptive learning design loop. Existing data from previously synthesized and characterized samples, together with relevant features describing materials' properties, serve to train machine learning frameworks that assess the relationships between independent descriptors (e.g., Fe loading, phenanthroline and ZIF-8 contents, pyrolysis temperature, etc.) and dependent outcomes (e.g., measured catalyst activity). The computational models are leveraged to predict the outcomes and their associated uncertainty for given input descriptor values. Integration of machine learning models with uncertainty quantification and global optimization methods enable design of experiment tools that suggest new samples to be tested, with the objectives of discovering materials with target properties (exploitation) and decreasing model uncertainty (exploration). The red star in the bottom right schematic plot highlights that we expect a greater improvement of the numerical algorithms if we experimentally study next sample labeled #3, due to its larger uncertainty, despite sample #2 showing an apparent higher value for the output variable. The results from new measurements augment the materials database, allowing for more robust computational models in the next iteration of the loop.

Figure 2

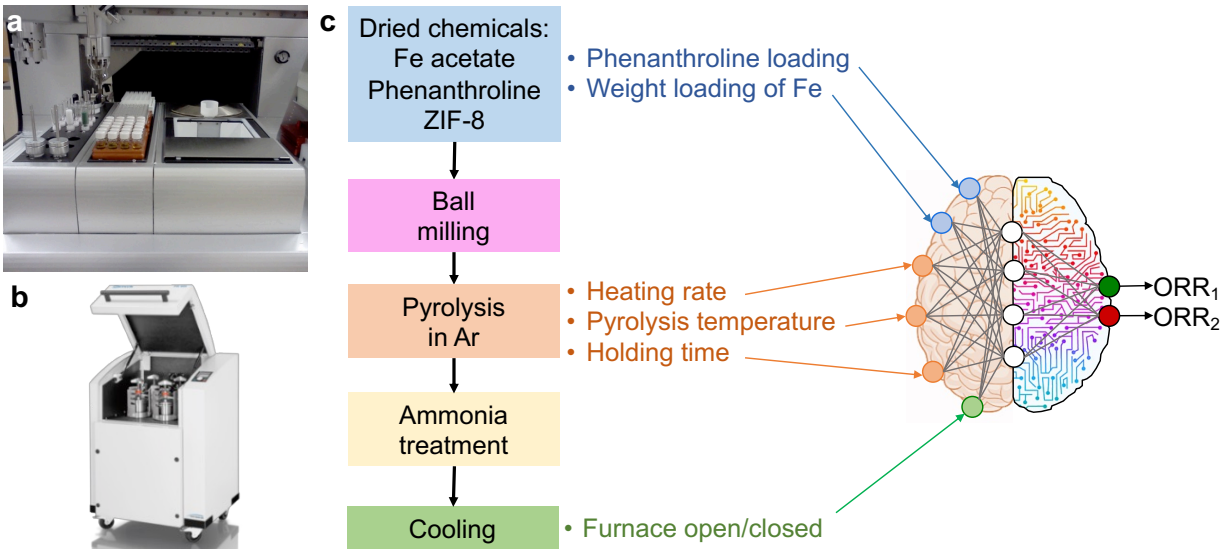

Fig. 2. High-throughput synthesis and machine learning integration. (a) Picture of the robotic synthesis platform (Junior, Unchained Labs) and (b) schematics of a high-throughput planetary ball mill (PM400, Retsch) available at Argonne National Laboratory's High-throughput Research Facility, which were used in this work for expedited synthesis of Fe-N-C electrocatalysts for oxygen reduction reaction. (c) Six independent parameters that can be controlled during the synthesis process have been selected for describing the catalyst samples, while the oxygen reaction rate mass activity measured in first and second cycles of staircase voltammetry in rotating-ring disk electrodes are used as dependent outcomes. The machine learning regression models surrogate the relations between the six input and two output features, allowing one to computationally elucidate the impact of synthesis conditions on catalyst activity and short-term stability.

Figure 3

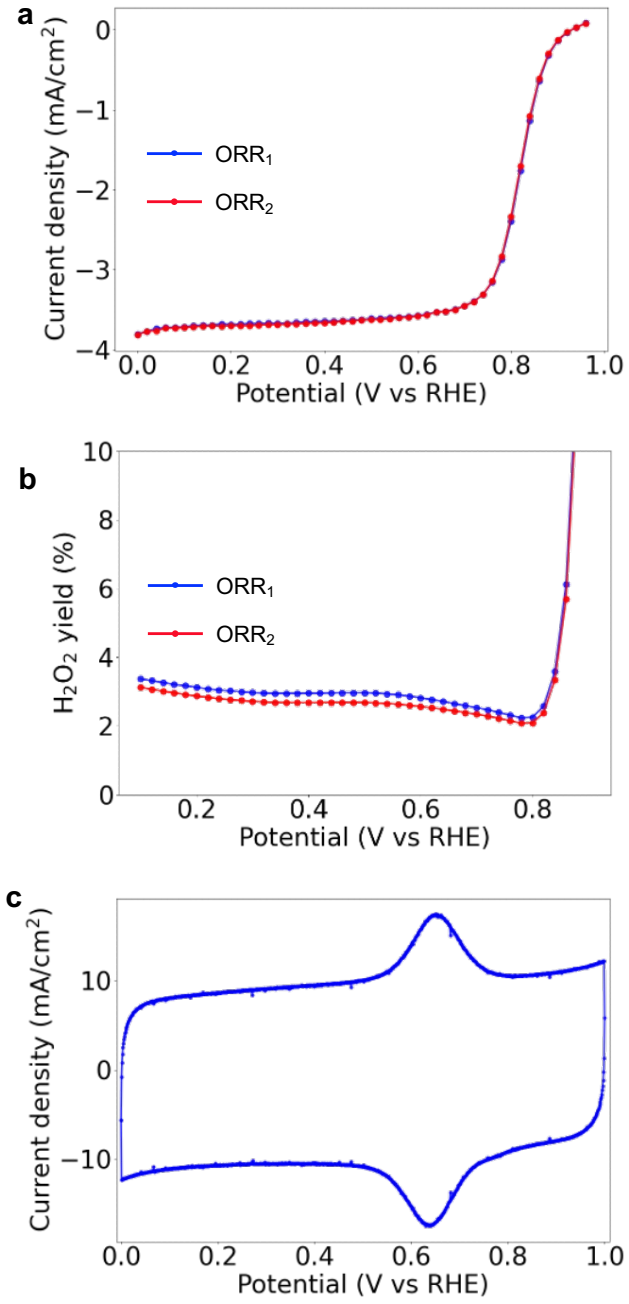

Fig. 3. Catalyst activity characterization. (a) Typical steady-state oxygen reduction reaction polarization curves at 900 rpm. (b) H_2O_2 yield as a function of disk electrode potential. (c) Representative cyclic voltammogram with 100 mV/s and 0 rpm in O_2 saturated 0.5 M H_2SO_4 (from sample #5 in the furnace open dataset).

Figure 4

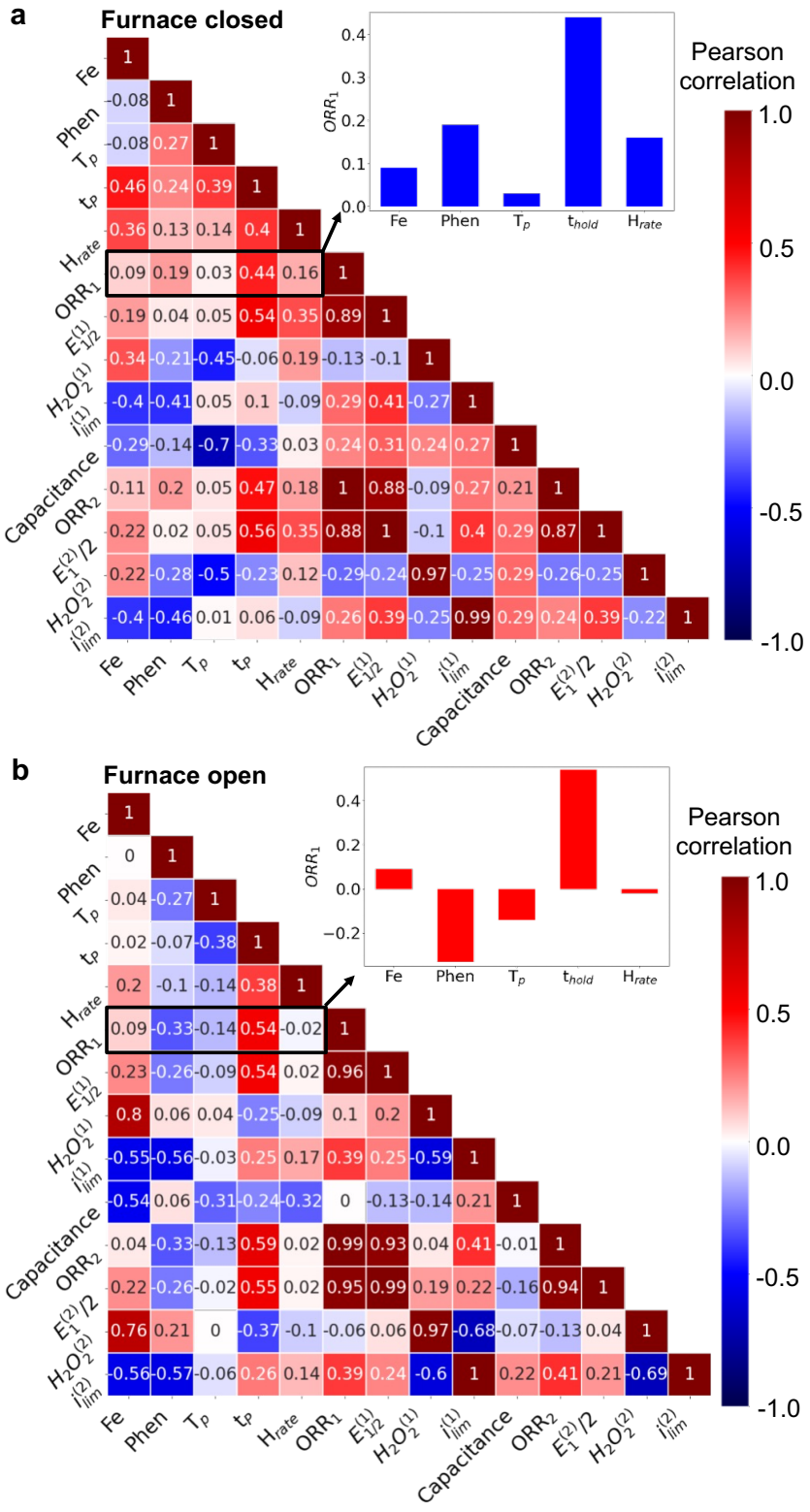
a
Furnace closed
Pearson correlation
1.0
0.5
0.0
-0.5
-1.0
Fe
Phen
T_p
t_p
H_{rate}
ORR_1
$E_{1/2}^{(1)}$
$H_2O_2^{(1)}$
$i_{lim}^{(1)}$
Capacitance
ORR_2
$E_1^{(2)}/2$
$H_2O_2^{(2)}$
$i_{lim}^{(2)}$
t_{hold}
b
Furnace open

Fig. 4. Feature correlations in synthesized electrocatalysts. Pearson correlation matrix for the two initial data sets with eighteen catalysts each, corresponding to samples cooled down to room temperature while keeping the furnace (a) closed or (b) open. The inset histograms show the linear correlations between the initial activity of the catalysts and the synthesis condition parameters. Correlations were computed for the six independently controlled features (i.e., Iron weight loading – Fe; phenanthroline-to-ZIF ratio – Phen; pyrolysis temperature – T_p; pyrolysis time – t_P; and heating rate – H_{rate}), the two dependent oxygen reduction reaction rate descriptors (ORR_1 and ORR_2), and various other measured properties of the samples (i.e., half-wave potential – $E_{1/2}^{(1)}$ and $E_{1/2}^{(2)}$; hydrogen peroxide yield at 0.4 V vs RHE – $H_2O_2^{(1)}$ and $H_2O_2^{(2)}$; limiting current – $i_{lim}^{(1)}$ and $i_{lim}^{(2)}$; and electrode capacitance). Superscripts in some of the variables indicate the corresponding voltammetry cycle of their measurement.

Figure 5

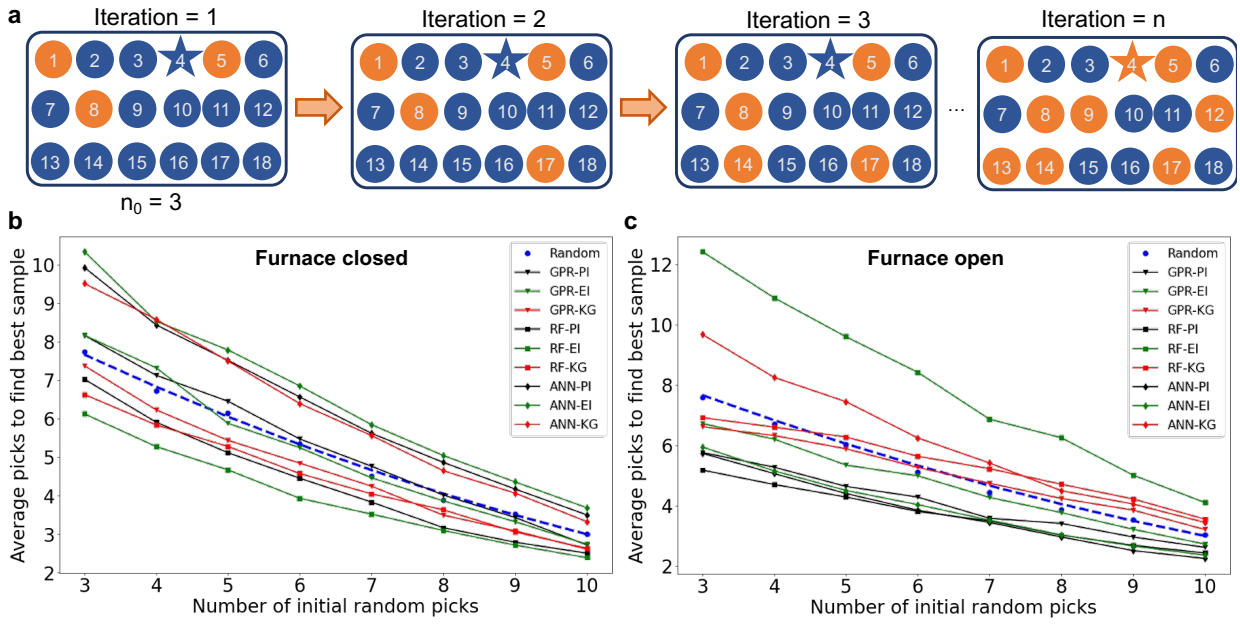

Fig. 5. Machine learning model-acquisition function optimization. (a) Schematic representation of the inference process developed for investigating the best regressor-selector pairs for the cases of furnace closed and open. n_0 samples are randomly picked from the initial data set and used to train a regression algorithm. An acquisition function is used to guide the next sample to be selected from within the remaining ones and it is then incorporated to the training data. The model is retrained with the $n_0 + 1$ samples, and the same global optimization method is used again to suggest the next sample to be picked. The process is repeated iteratively until the best sample (marked with a star, highest ORR_1 activity) in the initial database is discovered. The best machine learning model-acquisition function pair efficiently leverages the statistical information in the training dataset and finds the optimal synthesis conditions in as few iterations as possible. (b) Comparison of the performance of different combinations of machine learning models and acquisition functions on the original data set with a total of eighteen samples synthesized with the furnace closed. (c) Same as the previous panel for the case of the furnace open. In both (b) and (c) the blue dashed curve corresponds to random trial-and-error selection of electrocatalysts.

Figure 6

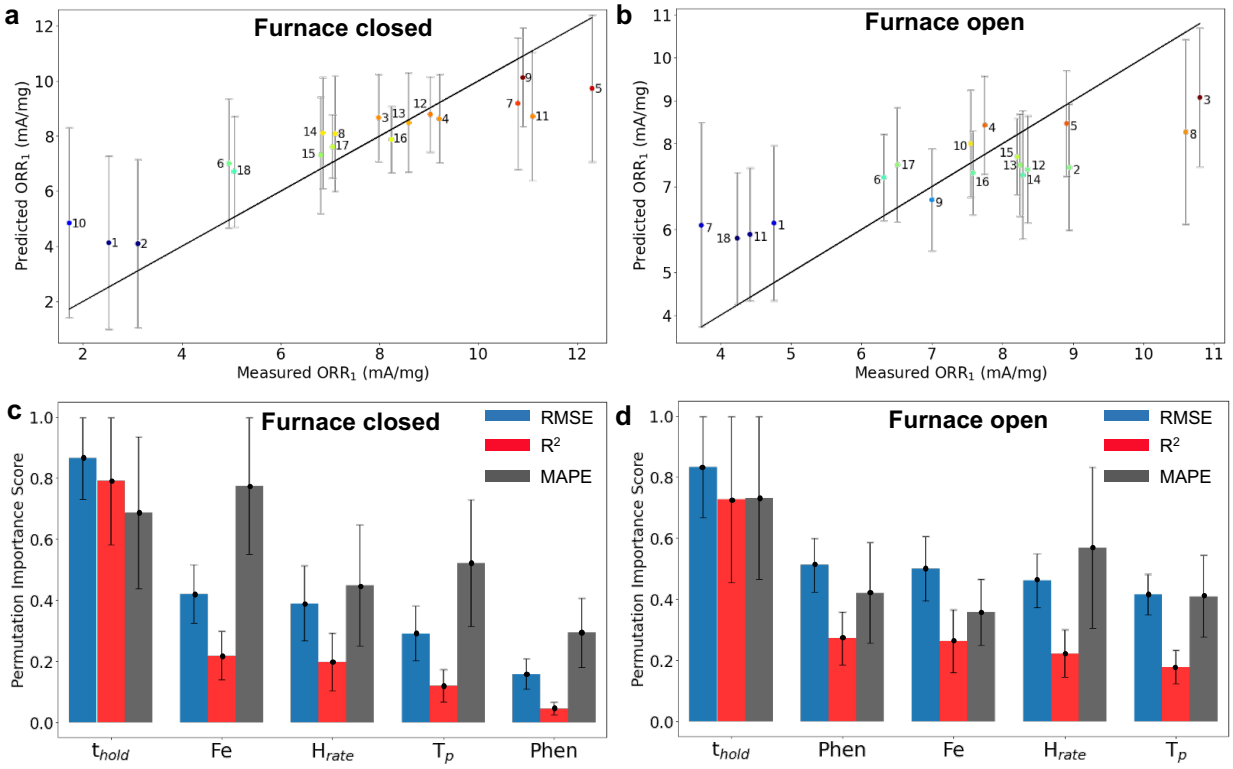

Fig. 6. Machine learning regression models predictive capabilities. Machine learning predicted oxygen reduction reaction activity versus actual measured values in the initial data sets for the furnace (a) closed and (b) open. Labels next to each point correspond to the order each sample appears in the databases in the supporting information. Following the results in Fig. 5, we have selected a random forest algorithm to model the catalyst activity for the furnace closed, and an artificial neural network model to surrogate the input-output correlations when the furnace is open. Computed values and associated errors are given by the means and standard deviations from predictions of machine learning models trained on 1,000 bootstraps (with replacement) of the original datasets. The black solid line represents ideal perfect agreement between model and measurements. Relative importance of the synthesis conditions descriptors on the catalyst activity for the furnace (c) closed and (d) open. The scoring metrics used here are the root-mean-square error (RMSE), the coefficient of determination (R^2), and the mean-absolute-percentage error (MAPE). Permutation importance score values and their uncertainties are averages and standard deviations, respectively, from 1,000 random shuffling operations in each feature.

Figure 7

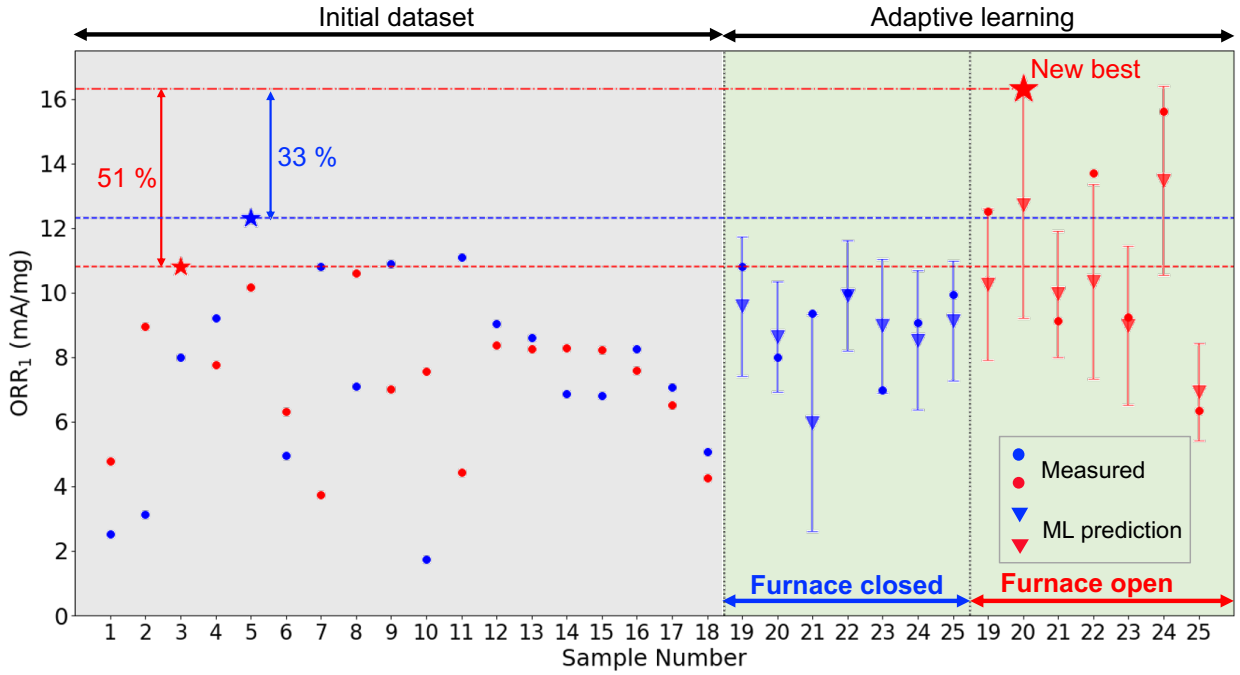

Fig. 7. Experimental results of adaptive learning predicted samples. Oxygen reduction reaction activity for the initial databases (1-18) and for the samples predicted via the developed design of experiment tool (19-25) for the furnace closed (blue markers) and open (red markers). See Tables S1-S4 for samples' synthesis conditions. Dot markers are experimental measurements while triangular markers correspond to predicted values from the machine learning models with uncertainties computed via statistical bootstrap sampling. Blue and red stars in the left-hand side show the highest ORR activity in the initial data for each furnace state. The red star on the right-hand side highlights the newly achieved best catalyst activity following the developed adaptive learning framework.

Figure 8

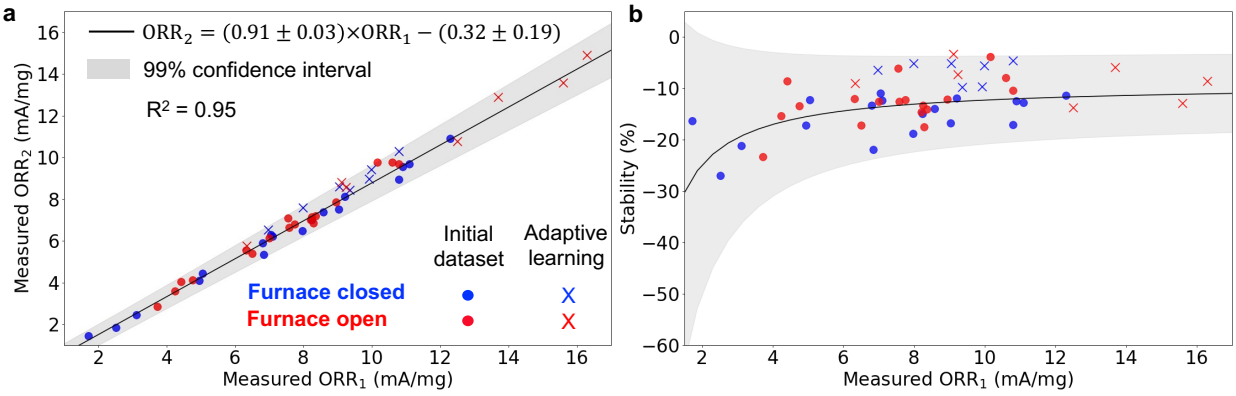

Fig. 8. Fe-N-C electrocatalyst short-term stability. (a) Measured ORR activity in the second RRDE voltammetry cycle (ORR_2) versus the initial activity (ORR_1) for the original (dot markers) and machine learning-guided ('x' markers) samples for the furnace closed (blue) and open (red). The solid black line corresponds to a linear regression model developed on the initial dataset considering all initial 36 samples (used as a training data set), regardless of the furnace state. The model generalizes well ($R^2 = 0.95$) for the 14 samples synthesized during the adaptive learning loop (used as a testing data set). The gray shaded area gives the 99% confidence interval of the model, computed via bootstrap sampling. (b) Short term stability of the synthesized samples, defined as $Stability\ (\%) = 100 \times (ORR_2/ORR_1 - 1)$, shows that highly active electrocatalysts ($ORR_1 > 8$ mA/mg) loose about $9 \pm 3\%$ of their initial activity between the first and second RRDE voltammetry cycles. The solid black line and the shaded area follow from the linear regression model depicted in panel (a).