ANNUAL REVIEWS

Annual Review of Condensed Matter Physics

Data-Driven Discovery of Robust Materials for Photocatalytic Energy Conversion

Arunima K. Singh,[1] Rachel Gorelik,[2] and Tathagata Biswas[1]

[1]Department of Physics, Arizona State University, Tempe, Arizona, USA; email: arunimasingh@asu.edu

[2]School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, Arizona, USA

Annu. Rev. Condens. Matter Phys. 2023. 14:237–59

First published as a Review in Advance on November 29, 2022

The *Annual Review of Condensed Matter Physics* is online at conmatphys.annualreviews.org

https://doi.org/10.1146/annurev-conmatphys-031620-100957

Keywords

photocatalysis, data-driven, materials discovery, aqueous stability, energy conversion, renewable energy

Abstract

The solar–to–chemical energy conversion of Earth-abundant resources like water or greenhouse gas pollutants like CO_2 promises an alternate energy source that is clean, renewable, and environmentally friendly. The eventual large-scale application of such photo-based energy conversion devices can be realized through the discovery of novel photocatalytic materials that are efficient, selective, and robust. In the past decade, the Materials Genome Initiative has led to a major leap in the development of materials databases, both computational and experimental. Hundreds of photocatalysts have recently been discovered for various chemical reactions, such as water splitting and carbon dioxide reduction, employing these databases and/or data informatics, machine learning, and high-throughput computational and experimental methods. In this article, we review these data-driven photocatalyst discoveries, emphasizing the methods and techniques developed in the last few years to determine the (photo)electrochemical stability of photocatalysts, leading to the discovery of photocatalysts that remain robust and durable under operational conditions.

1. INTRODUCTION

The photocatalytic solar–to–chemical energy conversion presents an alternative route to the generation of clean, renewable, and environmentally friendly fuels and chemical feedstocks. A photocatalytic process involves the absorption of sunlight by a photocatalytic material to generate electron–hole pairs. Unbound electron–hole pairs that diffuse to the surface of semiconducting photocatalytic materials can facilitate chemical reactions such as splitting of Earth-abundant water to hydrogen gas and oxygen gas. A photocatalytic material can also be used to convert pollutants such as carbon dioxide to useful chemicals and fuels, like alkanes, alcohols, and aldehydes.

Since the discovery of the first photocatalytic material, TiO_2, in 1972 by Fujishima & Honda (1), hundreds of photocatalysts have been identified (2–21). However, numerous challenges remain in the practical application of photocatalysis for industrial-scale fuel and chemical production. A techno-economic report by the Department of Energy (DOE) found 4–12% solar-to-hydrogen efficiency of photoelectrochemical hydrogen production in 2011, and the ultimate target of 25% remains unattained (22). The field of photocatalytic CO_2 reduction is even less well established. It is even more challenging to achieve high-efficiency solar-to-chemical energy conversion due to multiple competing reduction products, among other reasons.

Numerous remediation methods such as novel catalyst morphology, doping, alloying, use of cocatalysts, sacrificial agents, and promoters have been utilized successfully to improve the efficiency and selectivity of photocatalysts up to the current levels (23–25). In order to take the next leap in advancement of photocatalysis, there is a need for the discovery of photocatalysts that simultaneously exhibit long-term durability, high efficiency, and high selectivity toward their reaction products. A 2014 DOE report lists the discovery and development of a new generation of high-efficiency, high-durability photocatalytic materials and devices as one of the five major activities toward achieving their technical targets for competitive hydrogen production (22). In another report from 2019 (26), the DOE underscores the importance of identifying and understanding catalyst degradation pathways during CO_2 reduction reactions (CO2RRs) that occur in widely varying chemical environments, such as highly reducing conditions or extremely low or high pH. Most recent review articles highlight the importance of considering the photoelectrochemical stability of photocatalysts and recommend that future research efforts should aim to develop catalysts that are not only catalytically active and product-selective but also stable (27–34).

In the past century, materials discovery efforts have been largely trial and error based, but recently, large-scale data-driven discovery of materials has been undertaken by researchers, leading to the discovery of hundreds of new materials (12–21). These data-driven discoveries have been greatly accelerated due to the Materials Genome Initiative (MGI), launched in 2011, and the data revolution (35). The MGI has led to the generation, curation, and dissemination of high-quality computational as well as experimental data for materials structure and properties (36–43). It has also led to the use of data informatics and machine learning (ML) for the discovery, design, and optimization of catalysts.

In this article, we review recent advances in the data-driven discovery of photocatalyst materials, underscoring the importance of considering the durability and stability of catalysts under operational conditions. Recent breakthroughs in the methods and technique development for the assessment of electrochemical stability and their impact on the data-driven discovery of photocatalytic materials are discussed. Lastly, we discuss the future challenges and opportunities in the discovery of durable photocatalysts.

2. PHOTOCATALYTIC ENERGY CONVERSION PROCESS

In a photocatalytic energy conversion process, solar energy is converted to chemical energy. Incident sunlight of sufficient energy excites an electron across the optical band gap of a

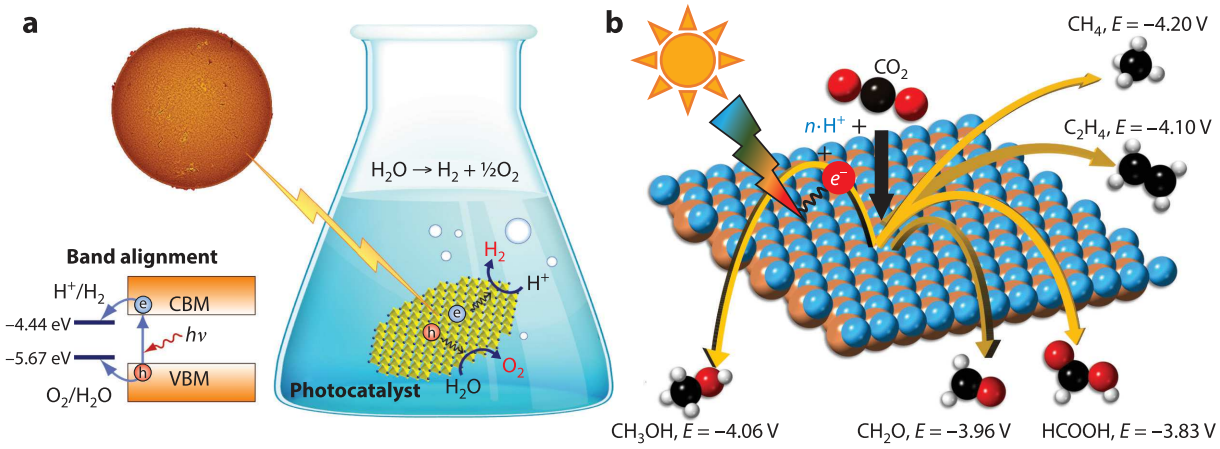

Figure 1

A schematic of the photocatalytic conversion of (*a*) water to hydrogen and oxygen gas (17) and (*b*) carbon dioxide to a number of reduced products such as alcohols, alkanes, and aldehydes (16). At a neutral pH, the potential required for conversion to each product is noted. Panel *a* adapted with permission from Reference 17, copyright 2015 American Chemical Society. Panel *b* adapted from Reference 16 (CC BY 4.0). Abbreviations: CBM, conduction band minimum; VBM, valence band maximum.

semiconducting photocatalytic material and generates an exciton, an electron–hole pair bound through Coulomb interaction. These excitons dissociate into free electron–hole pairs and diffuse to the solid–water interface, which can then drive the reduction or oxidation process of interest (e.g., in the water-splitting reaction or the CO2RR, as shown in **Figure 1**).

The photo-assisted splitting of Earth-abundant water leads to the production of "green" hydrogen gas. Hydrogen gas is a key twenty-first-century chemical—a clean fuel utilized in applications ranging from the production of ammonia to the oil-refining process. In a photocatalytic water-splitting reaction, the photoexcited electrons drive the hydrogen reduction reaction, generating hydrogen (Equation 1), whereas the holes participate in the oxidation reaction to generate oxygen (Equation 2) (17), as shown below:

$$2\,H^+ + 2\,e^- \rightarrow H_2, \qquad E^0 = 0.00\ \text{V}, \qquad 1.$$

$$H_2O + 2\,H^+ \rightarrow \frac{1}{2}\,O_2 + 2\,H^+, \qquad E^0 = 1.23\ \text{V}. \qquad 2.$$

Figure 1*a* shows a schematic of the water-splitting reaction enabled by a two-dimensional (2D) material. The overall reaction is nonspontaneous and requires a free energy of 1.23 eV. Thus, the band gap of photocatalysts must exceed this free energy of water splitting to show any activity. Aside from the suitable band gap, an ideal water-splitting photocatalyst must also have stability in water, the capability of capturing a significant fraction of solar energy, a conduction band minimum (CBM) energy higher than the reduction potential of H^+/H_2, and a valence band maximum (VBM) energy lower than the oxidation potential of H_2O/O_2. In addition, a low rate of backward reaction, low exciton binding energies, and low electron–hole recombination rates can further improve the efficiency of the photocatalytic water-splitting process. It is also desirable to use low-cost, abundant, and environmentally friendly materials for economical solar hydrogen production.

Photocatalytic solar–to–chemical energy conversion has been applied predominantly to the water-splitting reaction, but it has recently garnered attention for the reduction of carbon dioxide to fuels and chemical feedstocks (44). CO_2 is the primary greenhouse gas, accounting for 80%

of the anthropogenic greenhouse gas emissions in the United States (45). The conversion of CO_2 to fuels and feedstocks is an elegant solution to closing the carbon cycle when it is coupled with the use of a renewable energy source, such as in the case of photocatalysis. In the CO2RR, the photogenerated electrons react with carbon dioxide to generate products such as alcohols, alkanes, and formic acid (46), as shown below:

$$CO_2 + 2\,H^+ + 2\,e^- \rightarrow HCOOH, \qquad E = -0.61\,V, \qquad 3.$$

$$CO_2 + 2\,H^+ + 2\,e^- \rightarrow CO + H_2O, \qquad E = -0.53\,V, \qquad 4.$$

$$CO_2 + 4\,H^+ + 4\,e^- \rightarrow CH_2O + H_2O, \qquad E = -0.48\,V, \qquad 5.$$

$$CO_2 + 6\,H^+ + 6\,e^- \rightarrow CH_3OH + H_2O, \qquad E = -0.38\,V, \qquad 6.$$

$$CO_2 + 8\,H^+ + 8\,e^- \rightarrow CH_4 + 2\,H_2O, \qquad E = -0.24\,V. \qquad 7.$$

Figure 1*b* shows a schematic of the photocatalytic reduction of carbon dioxide to various reduction products. Like water-splitting photocatalysts, an ideal CO2RR photocatalyst must be stable in aqueous media, as water is used as a source of H^+. In addition, similar to the requirements for water-splitting photocatalysts, an efficient CO2RR photocatalyst must be capable of capturing a significant fraction of the solar energy, have a CBM higher in energy than the reduction potential, have low exciton binding energies, exhibit low electron–hole recombination rates, and exhibit a low rate of backward reaction. In contrast to the water-splitting reaction, the CO2RR can result in several competing products (Equations 3–7) or even the production of hydrogen gas. For an economical carbon dioxide reduction, it is necessary to have a high selectivity to just one product because the distillation of multiple products is expensive and energy consuming. Numerous experimental and theoretical studies have focused on exploring the large chemical space of materials to identify promising photocatalysts on the basis of both their intrinsic materials properties—such as thermodynamic stability, appropriate band gaps, and band edges—and their reaction-specific properties—such as the Gibbs free energy of absorption of reaction intermediates on the photocatalyst surface and the lowest energy barrier paths for the reactions. In particular, predictive modeling of structural, energetic, electronic, and surface properties has been possible through first-principles simulations, which have significantly advanced in the past decade. The concurrent increase in computing power, the development of materials databases, and the application of high-throughput simulation methods have allowed a data-driven approach to the discovery of photocatalytic materials. **Figure 2** shows an example of a data-driven approach to the discovery of water-splitting photocatalysts. However, the long-term robustness of photocatalysts, i.e., their electrochemical stability under operational conditions, has been far less addressed, partly owing to the challenges associated with the theoretical modeling and experimental measurements of the complex processes involved.

In the following section, Section 3, we highlight the necessity for assessing the robustness of photocatalysts. We also present an overview of recent successes in the development of theoretical methods for studying the robustness of photocatalysts, especially in a high-throughput manner.

3. ELECTROCHEMICAL STABILITY OF PHOTOCATALYSTS UNDER OPERATIONAL CONDITIONS

The electrochemical stability of a photocatalyst directly impacts its long-term durability under operational conditions. However, it is seldom evaluated in photocatalytic materials discovery studies, mainly owing to the challenges associated with the description of the temperature, applied potential, ionic concentration, and pH-dependent interactions at the solid–liquid interface of

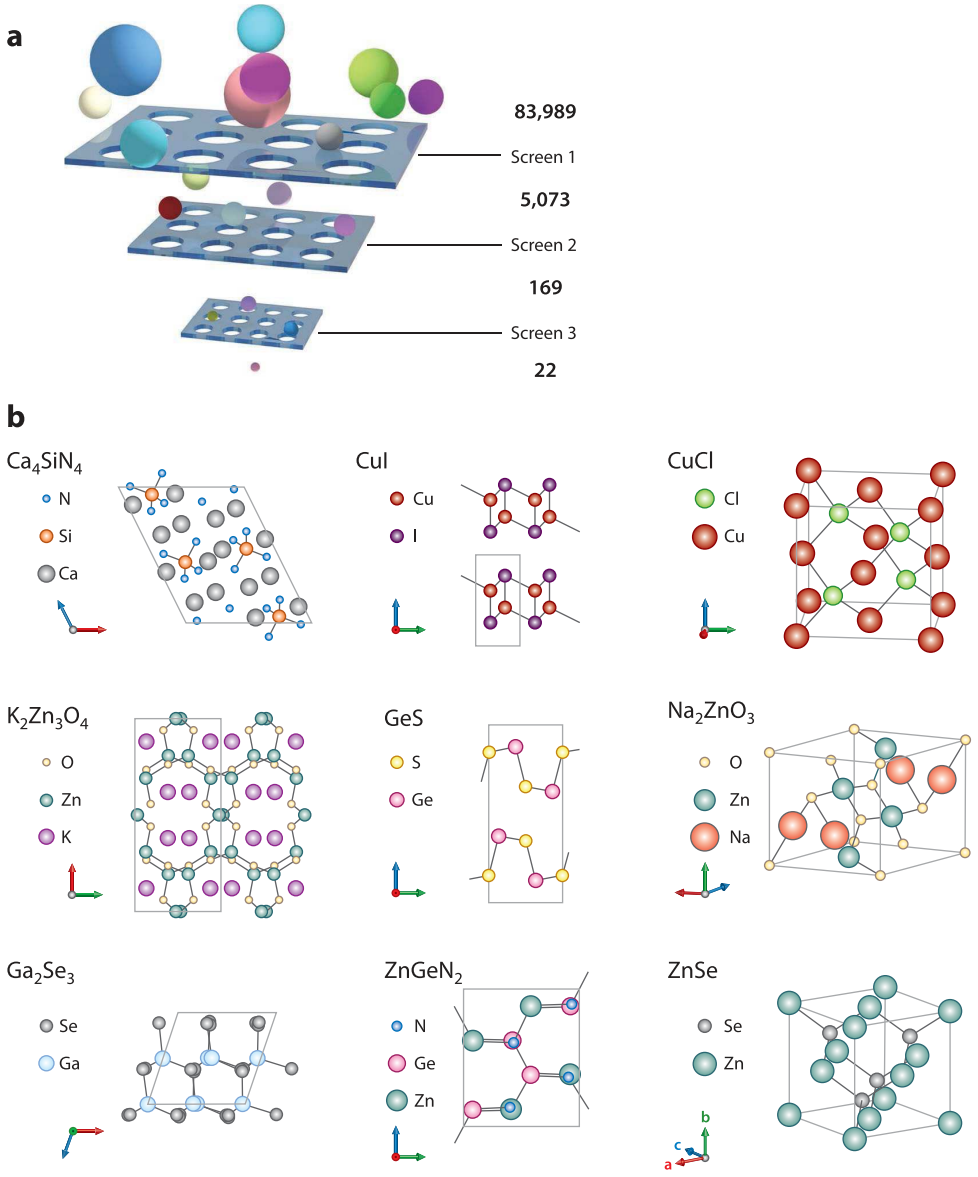

Figure 2

Jin et al. (47) performed a materials screening from 83,989 materials in the Materials Project database to discover water-splitting photocatalysts. (*a*) Some 5,073 materials emerge through formula and stability screens; of those, 169 materials satisfy the band gap and band edge position screens, and of those only 22 satisfy the effective mass screen. (*b*) In order to obtain higher-accuracy electronic and optical properties of the screened 22 materials, they use DFT with a hybrid density functional, HSE06, and find that only 9 materials finally meet the band gap criteria. Abbreviation: DFT, density functional theory. Figure adapted with permission from Reference 47, copyright 2019 American Chemical Society.

the photocatalytic material's surface. For instance, from Equation 2 we can see that the room-temperature oxygen evolution reaction (OER) occurs in oxidizing conditions, 1.23 eV, meaning most photocatalysts are prone to oxidation under these conditions. In contrast, the CO2RR occurs in reducing conditions, below −0.5 eV, under which most materials tend to either hydrolyze to their elemental forms or dissolve as aqueous ions (48). Furthermore, modeling and experiments have shown that an acidic environment is more desirable for high-efficiency fuel-forming devices but that most materials tend to aggressively corrode in acidic conditions (49).

In Sections 3.1 and 3.2, we briefly review the existing methods that can describe the physical and chemical effects pertaining to the interaction of an aqueous electrolyte with a photocatalyst. These methods vary considerably in level of detail, accuracy, computational expense, and utility in data-driven materials discovery studies. In Section 3.3, we describe in detail one of the most utilized methods for the data-driven discovery of robust (photo)catalysts.

3.1. Insights from Thermochemistry

Some thermochemistry models and principles can provide insight into the dissolution of materials in aqueous media. For instance, generally, materials with larger cation and anion sizes and/or small size differences remain stable in aqueous media (48). Also, multivalent materials tend to have strong lattice energies in comparison to monovalent materials, which renders them more likely to be stable in water in comparison to their monovalent counterparts (48). However, the Gibbs free energy of dissolution of compounds is dependent not only on the lattice enthalpy but also on hydration enthalpy, the entropy of dissolution, and the entropy of hydration (48). First-principles quantum mechanical modeling can be used to estimate these quantities, albeit with some assumptions and limitations. However, thermochemical models do not consider the effect of potential and pH, severely limiting their applicability to catalytic processes.

3.2. First-Principles Implicit and Explicit Solvation Methods

First-principles modeling methods that explicitly consider the electronic structure alongside the atomic structure of the complete electrochemical system—the material and the solvent—can be highly accurate. However, an explicit quantum mechanical consideration of the solvent–solid system is prohibitively time-consuming and resource-intensive. Several review articles (50, 51) present an overview of the first-principles methods for the study of electrochemical systems, underscoring the trade-offs among solvent description, the electric field description, and the computational accuracy achievable by them. These methods can be broadly classified into three categories: (*a*) explicit ab initio simulations in which both the solvent and the solid's electronic structure are computed, (*b*) implicit solvation models, where only the solid's electronic structure is considered and the solvent is described by a continuum model or (*c*) hybrid models, where quantum mechanical methods that consider the electronic structure are used together with molecular-mechanics models that describe only the atomic structure. **Figure 3** shows a schematic of the methods in order of decreasing accuracy and computational time of the simulations, where the fully explicit quantum calculation (highest accuracy, highest cost) corresponds to applying an explicit quantum mechanical treatment of the aqueous solvent, and the fully implicit (lowest accuracy, lowest cost) corresponds to modeling the solvent through an implicit polarizable continuum model. The three categories of modeling techniques described above have been used predominantly to study surface charged layers, surface energies, adsorption of molecules, reaction mechanisms, or kinetics, but have found limited application in data-driven studies due to their computational cost. There is a significant reduction in computational cost for implicit and

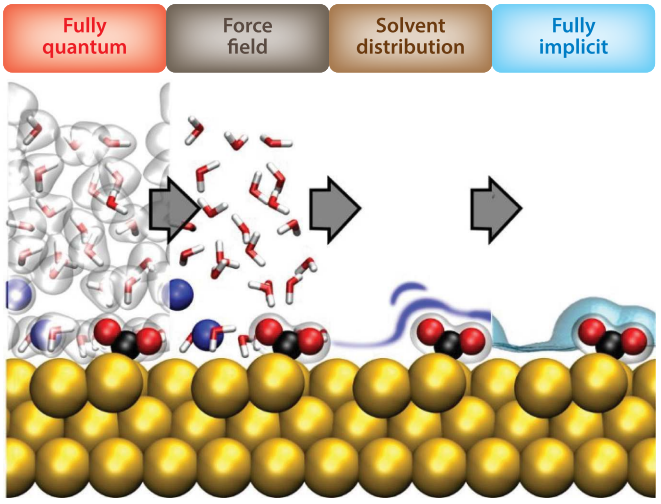

Figure 3

(*Left to right*) Methods for the modeling of the liquid phase in the context of catalysis at solid–liquid interfaces. The blue spheres denote the ionic species in the aqueous media, and the red and black spheres denote CO_2 molecules. The stick model denotes the water molecules, golden spheres denote the catalyst material, and isosurfaces indicate the electronic density of states. The catalyst surface is described on a quantum level throughout. The aqueous phase is described by fully explicit quantum mechanics (*far left*), by a force field or interatomic potential description (*center left*), only through their spatial distributions or correlation functions (*center right*), and finally by a fully implicit polarizable continuum model (*far right*). Figure adapted with permission from Reference 50, copyright 2021 American Chemical Society.

hybrid methods in comparison to the explicit methods, due to the elimination of the quantum mechanical treatment of the solvent, which also has very high configurational entropy. However, hybrid and implicit methods are still computationally expensive and, thus, are not pragmatic for any data-driven study.

3.3. Electrochemical Stability Maps

The data-driven discovery of robust materials has been most successfully treated by a first-principles-based formalism developed recently by Singh et al. (52). Singh et al.'s formalism combines first-principles simulation-computed energies to generate phase maps, called electrochemical stability maps (ESMs), in the Eh–pH space (i.e., plotted as a function of solution pH and applied potential; 53). This methodology enables the evaluation of the relative Gibbs free energy, ΔG_{pbx}, of stable and metastable materials as a function of pH, potential, temperature, and concentration of ionic species. Because the formalism relies on precomputed formation energies that are readily available in materials databases, it is orders of magnitude faster than the implicit, explicit, or hybrid solvation methods described in Section 3.2. This method is freely available on the Materials Project (MP) platform, in both a programmatic format and a web app, which enables large-scale electrochemical analysis of thousands of materials.

Because this formalism is universally applicable to all materials, including alloys and amorphous materials, it has been widely used for identifying robust materials for catalysis and photocatalysis. In fact, since its development in 2017, numerous studies that perform data-driven discovery of electrocatalysts (19, 27, 54–58) and photocatalysts (12, 16, 20, 21) have successfully employed this method, and many review articles highlight its usefulness (27–29, 31, 32, 34).

The instability landscape obtained from the ESMs provides a critical guide to environmental degradation management, as well as the design of metastable materials or protective films in materials of all classes for aqueous media–based applications.

3.3.1. Electrochemical stability map methodology. In this subsection, we describe the ESM formalism, with a particular emphasis on its applicability to data-driven catalyst discovery studies.

The ESMs of materials are based on the Eh–pH, or so-called Pourbaix diagrams, of the constituent elements of that particular material (52). Pourbaix diagrams give a concise equilibrium thermodynamics–based measure of the potential and pH ranges where the different multicomponent oxidation states of materials exist (53). Although the Pourbaix diagrams of elemental systems can be obtained using experiments (53), for multicomponent systems there exist an enormous number of possible redox reactions, making the experimental mapping of Pourbaix diagrams infeasible.

However, one can compute the Pourbaix diagram of a multicomponent system if all the possible equilibrium redox reactions as well as the standard state reaction Gibbs free energy of each redox reaction, $\Delta_r G^o$, are known. For example, a binary *M-X* system in aqueous media can result in a redox reaction such as

$$a\,M\,(\mathrm{s}) + b\,X\,(\mathrm{s}) + c\mathrm{H_2O}\,(\mathrm{l}) \rightarrow\ M_aX_b\mathrm{O}_c\mathrm{H}_d^m\,(\mathrm{aq}) + (2c-d)\,\mathrm{H}^+ + (2c-d+m)\,e^-, \qquad 8.$$

where a, b, c, and d are the stoichiometric coefficients of M, X, O, and H, respectively, and m is the charge on the aqueous species, $M_aX_b\mathrm{O}_c\mathrm{H}_d^m$, which can be positive, negative, or zero. The number of competing redox reactions increases rapidly with the number of elements for which the Pourbaix diagram is computed. At equilibrium, the Nernst equation can be used to relate the cell potential, E^o, to the reaction Gibbs free energy, $\Delta_r G$, for each possible redox reaction. For instance,

$$-\nu\,F\,E^o = \Delta_r\,G = \Delta_r G^o + R\,T\,\ln Q = \Delta_r\,G^o + R\,T\,\ln\left[\frac{(a_P)^p\cdot(a_{\mathrm{H}^+})^b}{(a_R)^r\cdot(a_{\mathrm{H_2O}})^w}\right], \qquad 9.$$

$$-\nu\,F\,E^o = \Delta_r\,G^o + 2.303\,R\,T\,\log\left[\frac{(a_P)^p}{(a_R)^r\cdot(a_{\mathrm{H_2O}})^w}\right] - 2.303\,b\,R\,T\,\mathrm{pH}, \qquad 10.$$

where T is the temperature, F is the Faraday constant, R is the ideal gas constant, $\mathrm{pH} = -\log(\mathrm{H}^+)$, $\nu = (2c-d+m)$ is the number of electrons, $a_R^r = a_{M(\mathrm{s})}^a.a_{X(\mathrm{s})}^b$ is the activity of the reactants, $a_P^p = a_{M_aX_bO_c\mathrm{H}_d^m}$ is the activity of the products, $a_{\mathrm{H_2O}}^w = a_{\mathrm{H_2O}}^c$ is the activity of water, and $a_{\mathrm{H}^+}^b = a_{\mathrm{H}^+}^{(2c-d)}$ is the activity of hydrogen ions for the reaction in Equation 8.

In principle, identifying the redox reaction that minimizes the difference between the contribution from the cell potential and the reaction Gibbs free energy, $\min(\Delta_r G + \nu\,F\,E^0)$, is sufficient to identify the range of pH and E where a particular solid or ionic species is stabilized over others for a given temperature, concentration of ions, and stoichiometry of elements under consideration. Persson et al. (59) combined density functional theory (DFT)–computed energies of solids and the Gibbs energy of experimentally available ions and aqueous species (scaled appropriately) to predict Pourbaix diagrams of n-component materials ($n = 1$ to ∞).

However, Pourbaix diagrams account only for materials that are at thermodynamic equilibrium and, thus, provide no insight into the electrochemical stability of metastable materials, despite metastable materials finding practical applications in many commercial applications and research such as steels (60), ceramics (61), photocatalysis (17), and optoelectronics (62). Furthermore, among the 29,902 unique bulk crystalline phases reported in the Inorganic Crystal Structure

Database (ICSD) (63), the first-principles-based, zero-temperature MP database computations predict that 50.5 ± 4% (15,097) of the experimentally synthesized structures are metastable solid-state phases (64).

To understand the electrochemical response of metastable materials, Singh et al. expanded the Pourbaix diagram formalism of Persson et al. to enable the evaluation of the Gibbs free energy, ΔG_{pbx}, of metastable as well as stable materials with respect to Pourbaix stable phases as a function of pH, potential, temperature, and concentration of aqueous species. An ESM of any material can be generated through this formalism.

Figure 4*a* shows the ESM for a metastable triclinic $FeVO_4$ phase computed for a temperature of 300 K, where the color bar shows the ΔG_{pbx} of $FeVO_4$ superimposed on the Pourbaix diagram for the Fe-V-O-H system. The figure shows that ΔG_{pbx} is less than 0.5 eV/atom in a narrow range of potentials, 0–1.5 V, in both acid and basic media. At higher potentials, the $FeVO_4$ is unstable with respect to the Pourbaix stable species (noted in the domains on the diagram) by more than 1 eV/atom. The number of phases that coexist (two for $FeVO_4$) in the domains of the Pourbaix diagram and their molar fractions are determined on the basis of the stoichiometry of the material for which the ESM is made, as well as the Gibbs phase rule. Thus, the ESM provides not only a prediction of possible decomposition products at a particular pH and potential but also a quantitative measure, ΔG_{pbx}, for the propensity of materials to be stable in water, either by inherent stability or through the development of a passivating film.

3.3.2. Quantitative predictions of electrochemical stability from electrochemical stability maps. In order to obtain a quantitative estimate of ΔG_{pbx} that can ensure that a material remains stable, either by inherent stability or through the development of a passivating film, Singh et al. compared the predicted ΔG_{pbx} with the electrochemical behavior of materials reported in the experimental literature. They found a clear correlation between the value of ΔG_{pbx} and the propensity of 20 materials [including GaP, $Zn(FeO_2)_2$, Si, WSe_2, GaAs, Fe_2O_3, WO_3, and TiO_2] reported in the experimental literature to be stable, passivated, or corroded.

Figure 4*b* shows the ΔG_{pbx} of those 20 photoanode materials at 1.5 V versus reversible hydrogen electrode (RHE) at pH values where direct or indirect experimental measurements of aqueous stability and/or surface film composition were available. Materials with ΔG_{pbx} lower than 0.5 eV/atom with respect to the Pourbaix stable phases have been reported to be stable against corrosion if they were predicted to dissociate to at least one Pourbaix stable solid species (**Figure 4*b***). Although some of the materials were inherently stable, some were stabilized by forming a surface film of a more stable solid. The experimentally inferred compositions of these films was in excellent agreement with those predicted by the ESMs (52). In **Figure 4*b***, materials with ΔG_{pbx} higher than 5.5 eV/atom are predicted to have Pourbaix stable phases as only ionic species and, hence, are expected to corrode, in perfect agreement with existing experimental literature.

In addition to comparing with the experimental data available in the literature, Singh et al. grew a metastable triclinic $FeVO_4$ phase to make a direct comparison with the ESMs. Their experimental measurements via X-ray fluorescence (XRF) (**Figure 4*c***), X-ray diffraction (XRD), and X-ray photoelectron spectroscopy (XPS) showed that the electrochemical response of $FeVO_4$ (**Figure 4*d***) was in excellent agreement with the computational predictions with respect to both the stability and the composition of the self-passivated layers, as applicable.

Thus, it can be surmised that the ESMs provide a quantitative estimate of the electrochemical stability of materials and can also be used to identify conditions under which arbitrary materials could be stable, corrode, or form passivation layers, along with their respective compositions.

3.3.3. Experimental validation of electrochemical stability map formalism. Experimental studies have utilized the ESM formalism to gain insight into the stability of catalysts and the

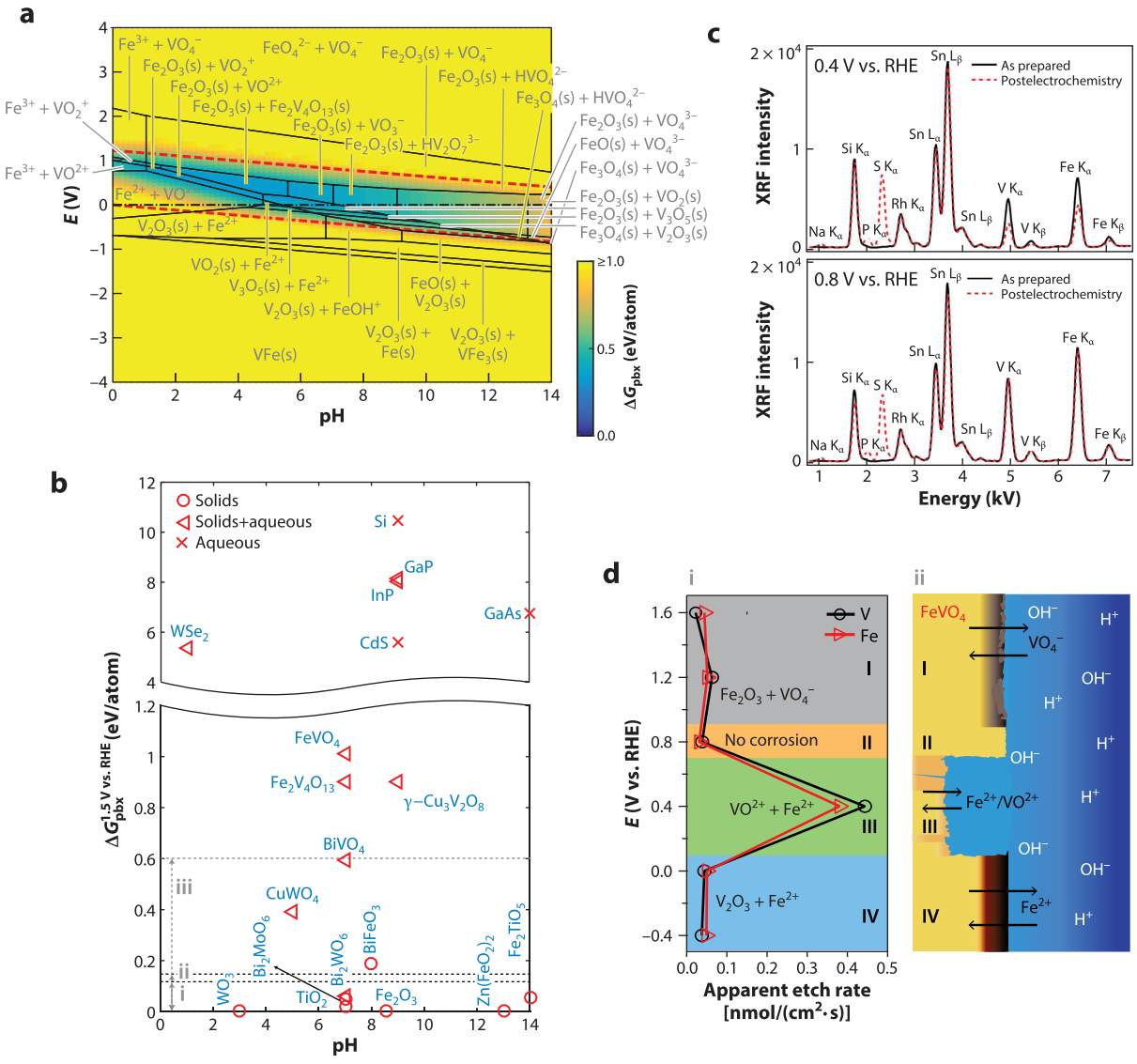

Figure 4

(*a*) Computationally predicted Pourbaix diagram of the Fe-V-O-H system. The Gibbs free energy, ΔG_{pbx}, of the metastable triclinic $FeVO_4$ phase with respect to the Pourbaix stable phases is superimposed and represented by the color bar. (*b*) ΔG_{pbx} of photoanode materials at 1.5 V versus RHE at pH values corresponding to those in experimental reports. Materials that are expected to remain stable or decompose to only solid species are shown as circles, those that decompose to solids as well as aqueous species are shown as triangles, and those that are expected to dissociate to only aqueous species are shown as cross symbols. (*c*) XRF spectra for an $FeVO_4$ sample before and after 15 min of an electrochemical stability measurement at pH 2.9 under the potential of 0.4 V and 0.8 V versus RHE. (*d*) (*i*) The apparent etching rate of Fe and V of the $FeVO_4$ sample over 15 min of aqueous electrochemical operations at pH 2.9 as a function of applied potential. The calculated stable species in the four pertinent voltage ranges, I–IV, are shown in background color. (*ii*) A schematic of the morphology of $FeVO_4$ in the presence of water under a steady-state flow of the majority ions through the solid–liquid interface in regions I–IV, corresponding to the voltage ranges marked in subpanel *i*. Abbreviations: RHE, reversible hydrogen electrode; XRF, X-ray fluorescence. Figure adapted with permission from Reference 52, copyright 2017 American Chemical Society.

composition of the potential passivating films that could be forming on catalysts (65–68). For instance, Wygant et al. (66) generated ESMs using the MP database for a variety of Ni *X*-ides (X = C, N, P, S, Se, and Te), OER catalysts that showed the speciation and stability of a selection of different catalysts across a range of pHs and potentials. They found that chalcogenides were generally much more stable than pnictides or carbides. However, all had a tendency to form oxide passivating films.

Hubert et al. (67) explored several Ru-based pyrochlores ($A_2Ru_2O_7$; A = Y, Nd, Gd, and Bi) as OER catalysts. Ru-based OER catalysts show significant promise for efficient water electrolysis, but rapid degradation poses a major challenge for commercial applications. They used in situ experiments, ex situ experiments, and ESMs to gain insight into the activity and stability of these pyrochlores under operational conditions. They thoroughly investigated the structure and composition of the catalyst during catalytic activity measurements and after catalytic activity testing. For instance, they were able to reliably quantify dissolution of >1% for the cationic species in the electrolyte by using inductively coupled plasma mass spectrometry. They also used ex situ XRD, scanning transmission electron microscopy (STEM)–energy dispersive X-ray spectroscopy, STEM–annular dark field, high-resolution transmission electron microscopy, scanning electron microscopy, Auger electron spectroscopy, and in situ X-ray absorption spectroscopy to characterize the catalyst surface. The ESMs from the MP database revealed that at OER-relevant potentials (~1.56 V versus RHE) in acidic conditions, the most stable states of the A-site species in the pyrochlores are aqueous ions, indicating that these species are likely to dissolve in the electrolyte. All of the $A_2Ru_2O_7$ structures considered in their study exceeded the ΔG_{pbx} empirical metastability threshold of 0.5 eV/atom under acidic OER conditions. The ESMs corroborated all of their experimental evidence of catalyst dissolution under acidic oxidizing conditions.

Stevens et al. (68) also did a rigorous study of electrocatalysts, focusing on their in situ characterization. Their material of focus for the oxygen reduction reaction (ORR) was MoN. Materials that contain light elements like N are known to be susceptible to both ex situ and in situ oxidation, leading to challenges associated with ex situ characterization methods. Utilizing electrochemical analysis, dissolution monitoring, and surface-sensitive X-ray techniques, they showed that under moderate polarization (0.3–0.7 V versus RHE) there is local ligand distortion, O incorporation, and amorphization of the MoN surface, without any changes in roughness. Furthermore, with a controlled potential hold procedure, they showed that the surface changes concurrent with potential conditioning were stable under ORR-relevant potentials. They found that at higher potentials (>0.8 V versus RHE), the films incorporated O, dissolved, and roughened, suggesting that in higher-potential regimes, the performance enhancements were due to increased access to active sites. They used ESMs to gain insight into the film stability and O incorporation as a function of potential. In agreement with the experiments, the ESMs predicted O incorporation in MoN at pH = 1 only at high potentials above ~0.5 V.

These experimental studies illustrate the use of the ESMs in predicting and understanding catalyst degradation and stability. They highlight the essential role that ex situ and in situ experimental characterization play in leveraging our understanding of surface dynamics to improve performance and design of catalytically active sites.

4. REVIEW OF DATA-DRIVEN DISCOVERY OF MATERIALS FOR PHOTOCATALYSIS

Recent advances in computing power, computational methods, and artificial intelligence have enabled the large-scale data-driven discovery of materials for catalysis. While many intrinsic properties of materials, such as their thermodynamic stability, band gaps, and band edges, have long

been considered in data-driven materials discoveries, the evaluation of robustness of the materials under operational conditions has been addressed in data-driven studies only in the last 2–3 years. In Section 4.1, we review the notable data-driven discovery efforts of photocatalytic materials. In Section 4.2, we highlight the data-driven materials discovery studies that also take into consideration the electrochemical stability of materials.

4.1. Photocatalyst Discovery Efforts That Do Not Consider Robustness of Materials

Some notable work has been done in the past few years regarding computational, joint computational and experimental, and ML-enabled discovery of photocatalysts (12–16, 20, 21, 27, 69–74). Some much earlier studies used high-throughput experimentation for the discovery of photocatalytic materials (75–77), but owing to the challenges associated with the time and cost of experimentation, the number of materials studied merely by high-throughput experimentation have been explored only within a limited phase space. In this section, we review the most notable and recent data-driven studies pertaining to discovery of photocatalysts.

The majority of studies reviewed in Sections 4.1 and 4.2 employ DFT simulations to compute the intrinsic properties of potential photocatalysts. DFT is a quantum mechanical method that is used to investigate the ground-state electronic structure of atoms, molecules, and crystals (78). In principle, DFT is an exact theory in which the total energy of a many-body interacting system is written as a functional of the electron density. However, the exchange correlation, E_{xc}, which is part of the total energy functional, remains unknown and is approximated at various levels of accuracy and complexity, such as through the local density approximation (LDA) or the generalized gradient approximation (GGA) (78). Ground-state properties such as lattice parameters, surface energies, and formation energies of materials can be predicted with high accuracy using DFT simulations.

Most often, in data-driven photocatalyst discovery studies, the DFT-computed formation energies of materials are used to assess the chemical stability of materials. Typically, the formation energy of a material is defined as the enthalpy change for the formation of one mole of a compound from its component elements. For example, let us consider the reaction $Al_2O_3 \rightarrow 2\ Al + \frac{3}{2}O_2$. We find that the DFT-computed formation energy of this reaction is −1,675.7 kJ/mol (79). Thus, 1,675.7 kJ of energy are released during the formation of 1 mole of Al_2O_3, making it more stable than aluminum metal and oxygen gas.

Although DFT is highly successful in predicting ground-state properties, the material properties that are an outcome of electronic excitations above the ground state, such as the band gap, are underestimated by the DFT method (78). The semiempirical DFT+U formalism provides a better prediction of band gaps and even the formation energies of materials (80). Hybrid functionals that include Hartree–Fock exchange in the E_{xc}, such as HSE or B3LYP, predict band gaps more accurately (81). Additionally, range-separated hybrid functionals like LRC-ω PBE and HSE06, which consist of a mix of short-range density functional exchange and long-range Hartree–Fock exchange, show further improvements in the accuracy of band gaps (82). Many-body perturbation theory simulation methods within the *GW* approximation provide an alternative approach to estimate band gaps and band edges of potential photocatalysts without the need for the adjustable parameters that are central to hybrid functionals (83). Although more predictive, the *GW* simulations are more resource intensive and time-consuming and, thus, are utilized only in a few data-driven photocatalyst discovery studies.

The band alignment of potential photocatalysts relative to the redox potentials of interest have also been predicted using the aforementioned computational techniques with the supercell slab method (16, 17). Similar to the estimation of band gaps, the accuracy and computational cost for the prediction of the band alignments are dependent on the computational method employed in

the study. Varying combinations of these methods as described above are therefore employed in the studies of photocatalyst data-driven discovery.

In a joint experimental and computational data-driven discovery effort, starting with 70,150 materials, Xiong et al. (14) short-listed 71 potential hydrogen evolution reaction (HER) photocatalysts by calculating their DFT-computed chemical stability, DFT+*U*-computed band gap, and DFT+*U*-computed band alignment. To avoid the computational requirement of large supercell calculations, the authors used the band gap and geometric mean of the Mulliken electronegativities of the constituent elements (84) to calculate the band alignment. **Figure 5*a*** shows the photocatalyst selection criteria used in their work; they started from a stability criterion and ended with phase purity evaluations using XRD, narrowing it to 11 candidates synthesized as single-phase materials. Six of those materials showed HER activity, but four of them corroded (Ca_2PbO_4, Ba_2PbO, $BaIn_2O_4$, and $BaCaFe_4O_8$).

Yan et al. (13) also did a joint experimental and computational screening to discover photocatalysts, but for OER. Their study was focused on the narrower phase space of the VO_4 ternary vanadates. Out of 174 known VO_4-based ternary vanadates in the MP database, they found through first-principles computations (DFT, DFT+*U*, and HSE) that only 34 materials satisfied their stability, band gap, and band edge criteria. Using combinatorial sputtering, they were able to grow 17 of them as high-purity materials. Some 15 of the synthesized vanadates exhibited OER photocurrent density.

Employing a computation-based approach to the discovery of photocatalysts, Sawada & Nakajima (18) performed a high-throughput computational screening of 29,160 perovskite oxynitrides and oxides to discover water-splitting photocatalysts. By applying a DFT-based thermodynamic-stability and a DFT- and HSE06-based electronic-properties screening procedure to the entire set of compounds, they identified 42 potential perovskite photocatalyst candidates for visible-light water splitting, including 34 newly proposed perovskites. Similarly, Castelli et al. (85) investigated up to 19,000 perovskite compounds using computations to discover water-splitting photocatalysts.

Jin et al. (47) also performed a similar computation-based high-throughput screening, starting from all the materials available in the MP database at that time (a total of 83,989). Again using DFT-computed stability and DFT- and HSE06-computed electronic-property-based screening criteria, they first narrowed their search down to 22 viable candidates; their screening process is depicted in **Figure 2*a***. For these remaining candidates, they employed higher-accuracy DFT band gap calculations and finally identified nine materials that showed promise as water-splitting photocatalysts (structures shown in **Figure 2*b***). For a comprehensive review of similar computations or joint computation- and experiment-based photocatalyst discovery efforts, see References 17 and 33 and references within.

ML methods are increasingly utilized alongside computations and experiments to further accelerate the discovery of photocatalysts. A wide range of ML-based regression and classification algorithms are used in these studies, such as the random forest classification algorithm, support vector regression, kernel ridge regression, and LASSO (least absolute shrinkage and selection operator). The availability of these algorithms, data preprocessing routines, and model evaluation algorithms in open-source Python libraries like scikit-learn have further accelerated the use of ML in data-driven discovery studies (86, 87).

In a 2019 review article, Masood et al. (88) presented an overview of the different categories of photocatalysis domain knowledge and proposed strategies for integrating that domain knowledge into ML frameworks. They noted that computational materials databases are an excellent source of first-principles-computed properties of materials. However, they also noted that an insufficient volume of reliable experimental data, incomplete details in reported literature, inconsistencies

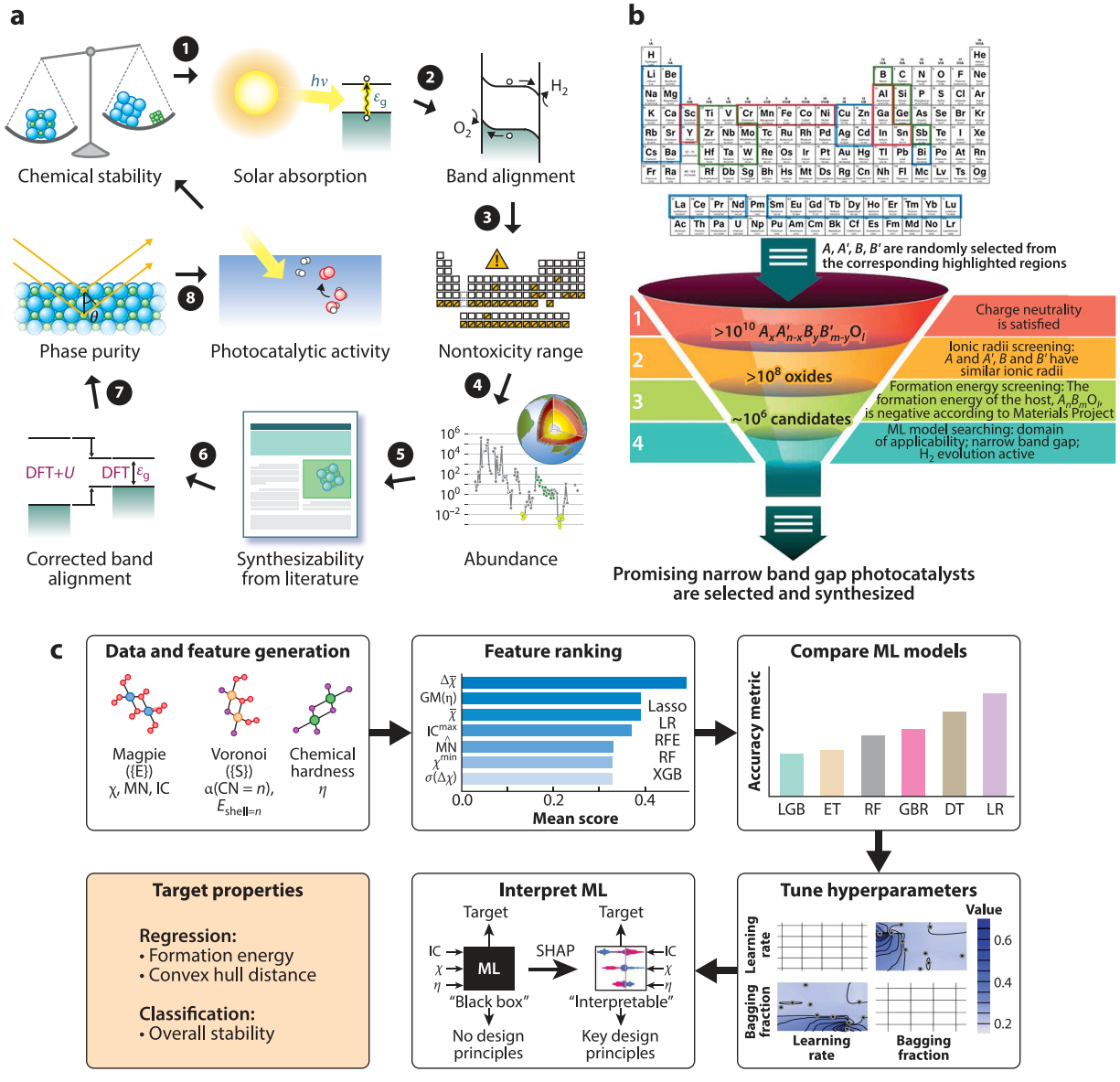

Figure 5

(*a*) A schematic of the joint computational and experimental photocatalyst selection strategy used by Xiong et al. (14) for the discovery of HER photocatalysts. Starting from 70,150 materials, they identified 11 potential photocatalysts, of which 6 showed appreciable photocurrent densities. (*b*) A library of $>10^{10} A_xA'_{n-x}B_yB'_{m-y}O_l$ materials was generated using ML models and screened for HER catalysts based on screening for charge neutrality, stability, ionic radii similarity, and photocatalytic activities (69). Some 20 novel nontoxic HER photocatalysts were predicted through this study. (*c*) An ML-based approach to the discovery of HER catalysts (70). Abbreviations: DT, decision tree; ET, extra tree; GBR, gradient-boosting regression; HER, hydrogen evolution reaction; IC, ionic character; LGB, light gradient boosting; LR, linear regression; ML, machine learning; RF, random forest; RFE, recursive feature elimination; SHAP, Shapley additive explanation; XGB, extreme gradient boosting. Panel *a* adapted with permission from the Royal Society of Chemistry from Reference 14; permission conveyed through Copyright Clearance Center, Inc. Panel *b* adapted from Reference 69 (CC BY 4.0). Panel *c* adapted from Reference 70 (CC BY 4.0).

in experimental measurement protocols, nonstandardized experimentation equipment, and challenges in acquiring new data (slow and expensive experiments and/or simulations) are the major hindrances to the use of ML for photocatalytic materials discovery.

Significant effort is underway to meet these challenges, as can be observed in the photocatalyst discovery-related articles published in the past year (15, 69–74). For instance, in a 2022 article, Mazheika et al. (73) reported that they generated a data set of CO_2 molecules adsorbed on 71 ternary and binary oxides, considering 141 surfaces. They applied a subgroup discovery algorithm to identify subgroups with outstanding properties (the C–O bond length, charge at the adsorbed CO_2, adsorption-induced dipole moment, etc.) that dictate suitability for CO_2 reduction. They used `RealKD` (**https://bitbucket.org/realKD/**) code for identifying one or more distinct combinations of material features (genes) that promote CO_2 activation. From these, they identified $CsNbO_3$, $CsVO_3$, $RbVO_3$, $LaScO_3$, $RbNbO_3$, and $NaSbO_3$ as the most promising candidates for CO2RR. Because CO_2 occurs in highly reducing conditions in which materials tend to hydrolyze to their elemental forms, it would be beneficial to assess the electrochemical stability of these materials for future experimental studies.

In another study, 14 potential perovskite water-splitting photocatalysts were screened from 30,000 ABO_3-type perovskite structures (71). Using ML models, the authors computed the materials screening criteria of structural stability, band gaps, conduction band energy, valence band energy, and hydrogen production rate. The ML tools used were `ExpMiner` (data mining software package) and `OCPMDM` (developed in-house) (89). They first built ML models from the known band gap and hydrogen production rate of ABO_3-type perovskite materials. The atomic parameters and experimental conditions were regarded as the variables for the ML model constructions. Then, 30,000 virtual samples of ABO_3-type perovskite structures were generated and filtered with the target values predicted by the established ML models. They identified 14 new materials through this approach. The average hydrogen production rate of these 14 perovskites was predicted to be 6.4% higher than the highest value in their training data set.

Mai et al. (69) utilized an experimental data set of 489 oxides to train regression models that predict band gaps and another data set of 380 oxides to develop hydrogen evolution classification models. They then created a library of more than 10^{10} compounds with the general formula $A_xA'_{n-x}B_yB'_{m-y}O_l$, where ($x > 0$, $x \geq n - x$, $y > 0$, $y \geq m - y$, and $l > 0$). **Figure 5*b*** shows that, after screening for charge neutrality, stability, and ionic radii similarity, they obtained a candidate data set of 10^6 compounds. They predicted the photocatalytic activities of this candidate data set using ML models. Using this work, promising photocatalysts predicted to have narrow bandgaps and HER activity were identified, and 20 novel nontoxic examples were identified.

In similar work, an experimental data set of 204 spinel structures with a chemical formula of AB_2X_4 (composed of 101 spinels from experimentally determined crystal structures and 103 spinels from DFT calculations) were used to train ML models (15). The models were then used to generate 4,160 different possible spinel structures, and eight spinels with direct band gaps and thermal stabilities at room temperature were screened out successfully.

Kumar & Singh (70) developed a database of 3,099 2D materials consisting of metals connected to six ligands in an octahedral geometry, termed the 2DO (octahedral 2D materials) database. ML models were then constructed using a combination of composition and chemical hardness–based features to gain insights into the thermodynamic and overall stabilities of the materials. **Figure 5*c*** shows their ML-based approach, which included feature ranking, selection of the best-performing ML algorithm, and tuning of ML hyperparameters in order to predict target properties for their screening. The most stable 2DO materials were further screened based on suitable band gaps within the visible region and band alignments with respect to standard redox potentials using the

GW method, resulting in 21 potential candidates. Furthermore, $HfSe_2$ and $ZrSe_2$ were found to have high solar-to-hydrogen efficiencies, reaching their theoretical limits.

In another study, the authors used an existing database to build generative models for the discovery of 2D photocatalysts for water splitting (72). Their generative models were used to produce several new 2D materials, and their band gaps were predicted through graph neural networks, whereas the band edge positions were computed via empirical correlations.

Thus, in the recent past, tremendous progress has been made in the integration of ML methods, computations, and experiments to enable large-scale data-driven discovery of photocatalysts. Nonetheless, many of these studies have yet to consider the electrochemical stability of the photocatalysts under operational conditions.

4.2. Photocatalyst Discovery Efforts That Consider Robustness of Materials

Just a handful of articles report the data-driven discovery of robust photocatalysts (12, 16, 20, 21, 27). However, many articles underscore the importance of determining the aqueous stability of (photo)catalysts, including several review articles (27–34). For instance, in a 2020 review article, Rajan et al. (28) noted that though the (photo)corrosion of photocatalysts has long been examined experimentally, such aspects have been much less addressed in the theoretical literature. Similarly, in another review article from 2020, Siahrostami et al. (27) highlighted the importance of evaluating not only the activity but also the electrochemical stability in the screening of $2e^-$ water oxidation materials. The inclusion of electrochemical stability is similarly crucial in high-throughput screening for photoanode materials, as Zhou et al. (29) underscored in their work on combinatorial screening of solar fuel photoanodes, stating that one of the major challenges in solar fuel photoanodes discovery is accounting for both catalytic activity and oxidative stability. Stein & Gregoire (31) discussed how databases, including computational Pourbaix stability databases, have increasingly become mainstays of data-driven research. Jain et al. (32) stated that Pourbaix diagrams and ESMs provide opportunities for the high-throughput discovery of materials. Pan & Yan (33) noted that the thermodynamic stability of photocatalysts has been addressed by constructing Pourbaix diagrams and ESMs but that there is a need for method development to allow the high-throughput study of the kinetic processes during photocorrosion. Holby et al. (34) stated that one major issue facing ORR photocatalysts is their insufficient long-term stability in environments where the ORR takes place.

4.2.1. Discovery of robust photocatalysts. In a purely computational and theoretical approach toward the discovery of photocatalysts, Singh et al. (16) performed a search for CO2RR photocatalysts, starting with 68,860 candidate materials. They identified 39 new photocatalytic materials that fulfill metrics for synthesizability and corrosion resistance under the highly reducing conditions (<-0.5 V versus RHE) needed for CO_2 reduction and which also exhibit band gaps and band edge energies suited for efficient solar energy conversion. The photocatalytic materials were identified via a systematic search based on first-principles simulations of intrinsic properties of the candidate materials, some available in public databases (DFT-based) and several computed specifically for their search through computationally expensive simulations (HSE06-based). Their computational strategy, shown in **Figure 6*a***, minimized the number of computationally expensive electronic structure simulations through a judicious screening of criteria (including thermodynamic stability, band gaps, and aqueous stability), which can be queried from existing information in databases (in their case, the MP database). Note that fewer than 5% of the semiconductors from tier 2 make it through the tier 3 criteria of aqueous stability, $\Delta G_{\mathrm{pbx}} < 0.2$ eV/atom, highlighting that very few semiconductors are water-stable at the reducing conditions needed for CO2RR. The 39 materials identified included arsenides, tellurides, selenides, and oxides, representing a wide

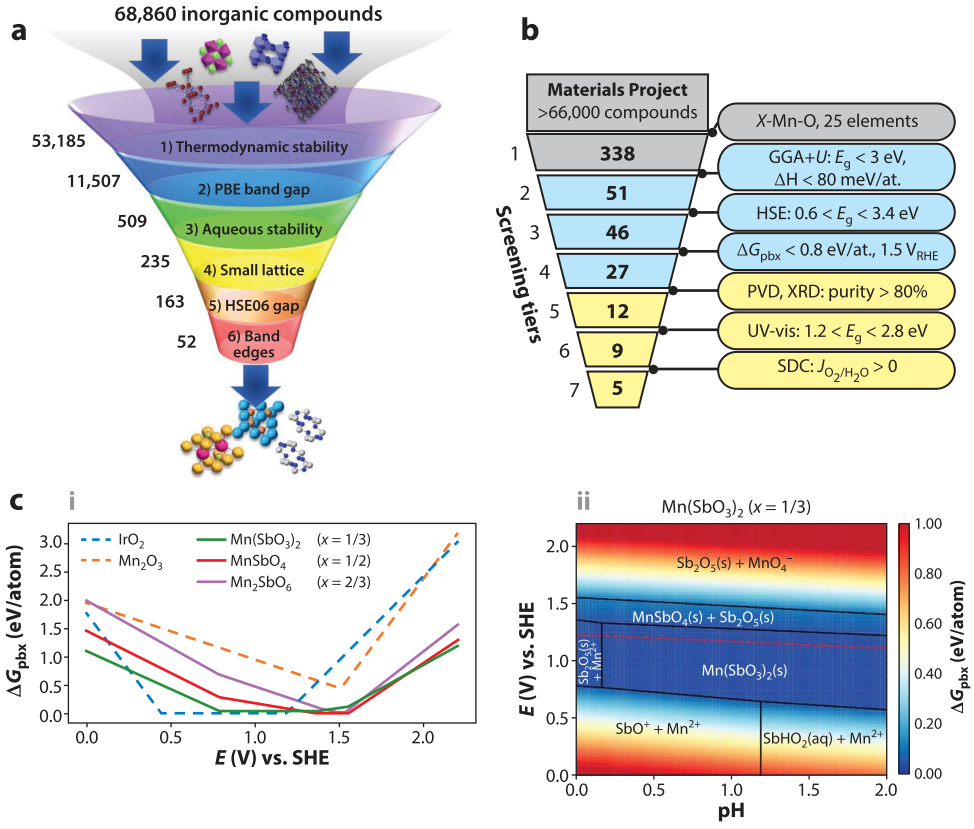

Figure 6

(*a*) The CO2RR photocatalyst materials selection criteria from Singh et al. (16). The number of materials that satisfy the criterion are shown for each tier. Note that fewer than 5% of the semiconductors from tier 2 make it through tier 3, highlighting that very few semiconductors are water-stable at the reducing conditions needed for CO2RR. (*b*) The screening tiers of the Shinde et al. (12) study, in which the authors used a joint computational and experimental screening approach to discover photocatalysts for OER. Quantities for the tiers marked in blue were determined from first-principles simulations, and those in yellow were from experimental measurements. (*c*) (*i*) The ΔG_{pbx} versus potential plot of various Mn-based oxides at pH = 0 (56) . (*ii*) The electrochemical stability map of $Mn(SbO_3)_2$. Panel *a* adapted from Reference 16 (CC BY 4.0). Panel *b* adapted with permission from Reference 12, copyright 2017 American Chemical Society. Panel *c* adapted with permission from Reference 56, copyright 2018 American Chemical Society. Abbreviations: PVD, physical vapor deposition; SDC, scanning droplet cell; SHE, standard hydrogen electrode; XRD, X-ray diffraction.

range of chemistries suited for extracting different CO2RR products. Biswas & Singh (90) further down-selected six materials based on the *GW* computed optical properties of these materials, taking excitonic effects into account via *GW*–BSE simulations.

Torrisi et al. (21) selected AlAs, AlSb, YbTe, ZnSe, ZnTe, GaSe, GaTe, and SiAs from the 39 photocatalysts that Singh et al. short-listed to identify monolayer counterparts of these materials and assess their efficiency for CO_2 reduction. They mapped these materials onto more than 200 prototype 2D materials structures available in the 2D materials database of Mounet et al. (91). After identifying the thermodynamically stable monolayer candidates from these structures, they verified the presence of visible-light band gaps, checked that band edges could support CO2RR, determined exciton binding energies, and computed surface reactivity. They used DFT, HSE06,

and *GW* methods for their study. They found visible-light absorption for SiAs, ZnTe, and ZnSe monolayers and that, owing to a lack of binding, CO selectivity is possible. Thus, they suggested SiAs, ZnTe, and ZnSe monolayers as targets for further investigation, expanding the chemical space for CO_2 photoreduction candidates.

Although computational and theoretical studies are able to perform data-driven discoveries of photocatalysts that start from hundreds and thousands of candidate materials, it is significantly more challenging to do so via experiments, as they are much more time-consuming and prohibitively expensive. To overcome this challenge, in a combined experimental and computational photocatalyst discovery effort, Shinde et al. (12) performed a discovery of Mn-based solar fuel photoanodes by conducting high-throughput experiments on candidates that were short-listed via thermodynamic stability, electronic structure calculations, and electrochemical stability modeling. In their work, 338 *X*-Mn-O phases were selected from the MP database, where *X* is one of the 25 elements that were chosen based on a number of criteria. They then down-selected 51 candidates based on their DFT-computed enthalpy, <80 meV/atom, and DFT+*U* band gap, <3 eV (properties which are available in the MP). In addition, they performed higher-accuracy DFT HSE band gap simulations and short-listed 48 candidates with band gaps 0.6–3.4 eV. Of these 48, only 27 phases were found to be stable over a broad pH 0–14 window by calculating ΔG_{pbx} at all 15 pH units in this range and requiring that the ΔG_{pbx} be below 0.8 eV/atom at a potential of 1.5 V. **Figure 6*b*** shows the screening tiers of this study, in which the quantities for tiers were determined from first-principles simulations and experimental measurements. Through high-throughput experiments, they synthesized multiple thin film libraries of 12 of these target phases with sufficient phase purity. They then did optical spectroscopy experiments and photoelectrochemistry in various electrolytes and found five phases that exhibited activity for OER.

In another combined experimental and computational photocatalyst discovery effort that assessed the electrochemical stability of materials, Zhou et al. (20) carried out combinatorial synthesis of 29 ternary vanadate and manganate oxides that had not previously been evaluated as photoanodes either experimentally or computationally, finding that 15 of them demonstrated visible-light response for oxygen evolution (20). Of those oxides, five were found to have photoresponses below 2.4 eV, including V_2CoO_6-trigonal and $Y_3Fe_5O_{12}$, which were therefore identified as the most promising photoanode materials from their search.

4.2.2. Discovery of robust electrocatalysts. In photocatalysis, solar energy is the source of electrons, whereas in electrocatalysis, the electrons are available from an applied bias. However, in both of these processes, a chemical reaction occurs on the catalyst surface with the aid of electrons. As both processes are closely related, and also for the sake of completeness, in this section we present a review of articles related to the data-driven discovery of robust electrocatalysts. The consideration of the electrochemical stability of materials is more prominent in the data-driven discovery of robust electrocatalysts (19, 27, 54–58, 92) than in that of photocatalysts.

Focusing on the Mn-Sb-O system in their discovery of acid-stable electrocatalyst materials for OER, Zhou et al. (56) considered the Mn-rich rutile alloy structures as promising candidates. They performed an in-depth experimental evaluation of not only their high catalytic activity but also their electrochemical stability at pH < 3, which they based on guidance from the generated ESMs [example map for $Mn(SbO_3)_2$ shown in **Figure 6*c***, subpanel *ii*]. From the ESMs, they were also able to construct plots of ΔG_{pbx} versus potential of different Mn-based oxides at pH = 0 to evaluate their aqueous stability compared to IrO_2 (as shown in **Figure 6*c***, subpanel *i*), seeing as the current acid-stable OER systems use extremely expensive iridium oxide catalysts. Through their work, they found the Mn-Sb-O materials to be economical alternatives to iridium oxide. Although

they found that their catalytic performance did not surpass that of IrO_2, the Mn-Sb rutile oxide alloys were predicted to exhibit higher robustness for OER in acidic conditions.

Jain et al. (19) also performed a study on the identification of acid-stable OER catalysts. They computed the electrochemical stability of 2D materials for the ORR and OER at the low pH value of 1. They tested the electrochemical stability of 11,000 2D materials that were available through 2D materials databases using the ESM formalism of Singh et al. At a pH of 1 and oxidizing potentials of 0.9 and 1.5 V for ORR and OER, respectively, and a criterion of the predicted $\Delta G_{pbx} < 0.2$ eV/atom, only 130 and 69 materials were found to be potentially stable under the strong oxidizing conditions of ORR and OER in the acidic medium, respectively.

In another high-throughput search for acid-stable catalysts, Wang et al. (58) explored the aqueous stability of 47,814 nonbinary metal oxides under typical oxygen reduction and evolution reaction conditions and found that only 68 were likely acid-stable candidates for oxygen electrocatalysts. These 68 materials had $\Delta G_{pbx} < 0.5$ eV/atom. They observed that Sb/Ti/Sn/Ge/Mo/W-based oxides have a preference for being stable in strong acids. Gunasooriya & Nørskov (57) then studied the adsorption of reaction intermediates and computed the surface effects on the Pourbaix diagrams for the unique surfaces of these ternary and quaternary materials identified by Wang et al. Based on their study, they identified $Co(SbO_3)_2$, $CoSbO_4$, $Ni(SbO_3)_2$, $Fe(SbO_3)_2$, $FeSbO_4$, $FeAg(MoO_4)_2$, $MoWO_6$, and $Ti(WO_4)_2$ as promising materials for OER performance.

In another work, Back et al. (92) performed a high-throughput computational data-driven study for the discovery of acid-stable, cost-effective, and active catalysts for OER. They performed first-principles-based high-throughput catalyst screening to discover OER-active and acid-stable catalysts focusing on equimolar bimetallic oxide crystal structures that were formed from eight crystal structure prototypes derived from IrO_x and TiO_x polymorphs and bimetallic combinations of 26 transition-metal elements (2,600 in total). Based on the computational results on acid stability and OER catalytic activity, they suggested a few Ir-based equimolar bimetallic oxides, including CoIr, FeIr, and MoIr combinations, which satisfied all the target properties and are predicted to outperform monometallic Ir oxides.

Back et al. (54) did another high-throughput computational screening to develop O_2 reduction catalysts for selective H_2O_2 production. They collected 3,569 binary alloys consisting of 48 elements from the MP database. They then short-listed 942 of these materials based on the condition that they should have both inactive and active elements (Ag, Au, Hg, S, Tl, Se, Br, C, Cl, F, I, and N) (54). Using the ESM formalism, they found that only 73 materials were stable with a ΔG_{pbx} of less than 0.5 eV/atom at three different pHs, i.e., $U(V) = 0.7–0.0591$ pH (pH = 0, 7, and 14) values. They then computed the O* and OOH* adsorption energies on various low-index surfaces of these materials and found that the surfaces of 22 materials had a low overpotential of 0.20 V and selectivity to H_2O_2 generation of <0.85 eV.

Another study in which hundreds of candidate materials were screened on the basis of first-principles-simulated intrinsic properties and their electrochemical stability was that by Siahrostami et al. (27). They reduced a pool of 559 perovskite materials to 59 based on thermodynamic descriptors of OH* and O* adsorption energies. By further applying the ESM formalism, they additionally filtered their search down to 10 perovskite structures that were predicted to be stable under $2e^-$ water oxidation reaction conditions. None of the perovskites were predicted to be stable under acid-harsh pH = 0 conditions.

In all of the studies reviewed in Section 4.2, we can clearly see that there is a significant reduction in the number of potential catalyst candidates upon determining the electrochemical stability of the materials. This spotlights the importance of determining aqueous stability of materials in the discovery of next-generation catalysts.

5. FUTURE DIRECTIONS

The ESM formalism and Pourbaix diagrams have clearly enabled a new wave in the data-driven discovery of robust photocatalysts in the recent past. Although validation with experiments shows that ESMs are quite predictive and convenient to generate, they are only able to provide insight into the initial and final states of a material under a given pH, potential, temperature, and concentration of ionic species.

One major limitation of the ESM formalism is that it does not give any information about the time evolution of the electrochemical process. Although ESMs can provide guidance for the propensity of materials to corrode, passivate, or be inert, they are unable to provide time-dependent properties like corrosion rate, passivating film growth rate, passivating film thickness, and protectiveness of the passivating film. This is of great consequence in practical applications. For instance, the ESMs may predict corrosion for a material under a certain condition, but it is possible that the material will have unnoticeable surface corrosion due to very low corrosion rates. Although first-principles methods, both implicit and explicit models, can be used to obtain an accurate description of some of the processes that dictate the kinetics, such as reaction energy barriers, these methods are not pragmatic enough to guide data-driven photocatalysts discoveries owing to their high computational cost.

An alternative method to modeling the kinetics of corrosion and passivation in a computationally tractable way could be to utilize first-principles-computed parameters in phenomenological well-established kinetic theories (93), such as the Wagner theory, the Jost–Hard–Price model, the point defect model, and the Gulbransen theory. The ESM formalism also lacks the description of surface effects like the surface energy of the photocatalysts, surface defects, adsorption of species present in the aqueous media, the interaction of the photogenerated electrons and holes with the adsorbed and ionic species, and the recombination of the photogenerated electrons and holes.

Surface Pourbaix diagrams can be computed to take into account the effects from the surfaces and adsorbate interaction energies, such as those computed by Hansen et al. (94) for Pt, Ag, and Ni (111) surfaces. The larger-scale study of various surfaces, the various concentrations of adsorbates on these surfaces, and the different adsorption sites on the surfaces is certainly possible due to the workflow codes developed in the recent past (95–97); however, there is still a need to curate these properties in a systematic manner in open-source databases to enable the discovery of electrochemically robust photocatalysts.

The ESMs that can be generated using the MP website and/or code base utilize the energies of solids computed with the DFT method using the PAW-PBE (projector augmented wave Perdew–Burke–Ernzerhof) pseudopotentials and the DFT+U method for some of the transition metal oxides. Wang et al. (98) have shown that using the strongly constrained and appropriately normed (SCAN) functional instead can improve the accuracy of predicted aqueous stability in a large number of cases of interest. In another study, Huang & Rondinelli (99) studied the effect of magnetism and hybrid functionals on the improvements to the Pourbaix diagrams of Cr-, Mn-, Fe-, Co-, and Ni-containing materials. In the future, the databases of energies computed from various advanced functionals, such as the SCAN functional and hybrid functionals, could enable the large-scale assessment of the functional-dependent electrochemical stability of materials.

There is also a need for thorough in situ and ex situ characterization of photocatalytic materials for the identification of corrosion mechanisms, development of next-generation theories of corrosion and passivation, and design of functional passivating films for photocatalysis. The experimental characterization of photocatalysts for signs of electrochemical degradation are usually inferred from reductions in photocurrent densities. However, a detailed in situ and ex situ characterization is needed to understand the thermodynamics and kinetics of the photocorrosion processes. In recent work using an experimental data set of organic molecules that induced

degradation of photocatalysts, Ayodele et al. (100) encoded features of the photocatalysts and the organic molecules as an input to an artificial neural network (ANN). This ANN not only achieved a very good prediction of the photocatalytic degradation rate constants by different photocatalysts over a wide range of organic contaminants but also gave insight into the factors that influence the photodegradation process. For example, the authors showed that the type of photocatalyst and its size had a relatively high impact on its photodegradation performance, whereas the initial concentration of the contaminants and the pH did not have as much influence on the photodegradation performance. Thus, combined experimental and theoretical investigation of the surfaces of the photocatalyst can provide much-needed advancement in photocorrosion control, improvement in design, and enhancement of durability of the photocatalyst.

6. SUMMARY AND OUTLOOK

In this review article, we have presented the current state of progress in the data-driven discovery of photocatalysts for solar–to–chemical energy conversion. We demonstrate that current studies have employed quantum mechanical simulations, atomistic simulations, ML, and high-throughput experimentation to discover hundreds of photocatalysts that are catalytically active, efficient, and selective. However, most of the data-driven discovery work does not consider the robustness of the materials under electrochemical operational conditions. We review the current methods to address the electrochemical stability of photocatalysts and highlight the significant progress in the discovery of robust photocatalysts that was enabled through the ESMs formalism. We present numerous future directions to further develop models and theories that can allow the discovery of robust photocatalysts for solar–to–chemical energy conversion.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation DMR grant 1906030, Arizona State University start-up funds, and the National Science Foundation Graduate Research Fellowship Program under grant 026257-001.

LITERATURE CITED

1. Fujishima A, Honda K. 1972. *Nature* 238(5358):37–38
2. Katz JE, Gingrich TR, Santori EA, Lewis NS. 2009. *Energy Environ. Sci.* 2:103–12
3. Sivula K, Van De Krol R. 2016. *Nat. Rev. Mater.* 1(2):15010
4. Cui Z, Zeng D, Tang T, Liu J, Xie C. 2010. *J. Hazard. Mater.* 183(1–3):211–17
5. Maschmeyer T, Che M. 2010. *Angew. Chem. Int. Ed.* 49(9):1536–39
6. Zhao ZG, Miyauchi M. 2008. *Angew. Chem. Int. Ed.* 47(37):7051–55
7. Zhou L, Wang W, Xu H, Sun S, Shang M. 2009. *Chem. Eur. J.* 15(7):1776–82
8. Ikeda M, Kusumoto Y, Somekawa S, Ngweniform P, Ahmmad B. 2006. *J. Photochem. Photobiol. A* 184(3):306–12
9. Jaramillo TF, Baeck SH, Kleiman-Shwarsctein A, McFarland EW. 2004. *Macromol. Rapid Commun.* 25(1):297–301
10. Qiu Y, Yang M, Fan H, Zuo Y, Shao Y, et al. 2011. *CrystEngComm* 13(6):1843–50
11. Habisreutinger SN, Schmidt-Mende L, Stolarczyk JK. 2013. *Angew. Chem. Int. Ed.* 52(29):7372–408
12. Shinde A, Suram SK, Yan Q, Zhou L, Singh AK, et al. 2017. *ACS Energy Lett.* 2(10):2307–12

13. Yan Q, Yu J, Suram SK, Zhou L, Shinde A, et al. 2017. *PNAS* 114(12):3040–43
14. Xiong Y, Campbell QT, Fanghanel J, Badding CK, Wang H, et al. 2021. *Energy Environ. Sci.* 14(4):2335–48
15. Wang Z, Zhang H, Li J. 2021. *Nano Energy* 81:105655
16. Singh AK, Montoya JH, Gregoire JM, Persson KA. 2019. *Nat. Commun.* 10:443
17. Singh AK, Mathew K, Zhuang HL, Hennig RG. 2015. *J. Phys. Chem. Lett.* 6(6):1087–98
18. Sawada K, Nakajima T. 2018. *APL Mater.* 6:101103
19. Jain A, Wang Z, Nørskov JK. 2019. *ACS Energy Lett.* 4(6):1410–11
20. Zhou L, Shinde A, Guevarra D, Richter MH, Stein HS, et al. 2020. *J. Mater. Chem. A* 8(8):4239–43
21. Torrisi SB, Singh AK, Montoya JH, Biswas T, Persson KA. 2020. *npj 2D Mater. Appl.* 4:24
22. Hydrog. Fuel Cell Technol. Off. 2014. *Multi-year research, development, and demonstration plan*. Res. Rep., Off. Energy Effic. Renew. Energy, Washington, DC. **https://www.energy.gov/eere/fuelcells/downloads/hydrogen-and-fuel-cell-technologies-office-multi-year-research-development**
23. Zhou M, Lou XW, Xie Y. 2013. *Nano Today* 8(6):598–618
24. Abe R. 2010. *J. Photochem. Photobiol. C* 11(4):179–209
25. Ni M, Leung MKH, Leung DYC, Sumathy K. 2007. *Renew. Sustain. Energy Rev.* 11(3):401–25
26. Basic Energy Sci. Roundtable Liq. Solar Fuels Panel. 2019. *Report of the Basic Energy Sciences Roundtable on Liquid Solar Fuels*. Final Rep., Off. Sci. Basic Energy Sci., US Dep. Energy, Washington, DC. **https://science.osti.gov/-/media/bes/pdf/reports/2020/Liquid_Solar_Fuels_Report.pdf?la=en&hash=06D037C1887D2FF8B872035E4C51FFDDEC11D4C8**
27. Siahrostami S, Villegas SJ, Bagherzadeh Mostaghimi AH, Back S, Farimani AB, et al. 2020. *ACS Catal.* 10(14):7495–511
28. Rajan AG, Martirez JMP, Carter EA. 2020. *ACS Catal.* 10(19):11177–234
29. Zhou L, Shinde A, Guevarra D, Haber JA, Persson KA, et al. 2020. *ACS Energy Lett.* 5(5):1413–21
30. Hoye RL, Schulz P, Schelhas LT, Holder AM, Stone KH, et al. 2017. *Chem. Mater.* 29(5):1964–88
31. Stein HS, Gregoire JM. 2019. *Chem. Sci.* 10(42):9640–49
32. Jain A, Montoya J, Dwaraknath S, Zimmermann NE, Dagdelen J, et al. 2020. In *Handbook of Materials Modeling: Methods, Theory and Modeling*, ed. W Andreoni, S Yip, pp. 1751–84. Cham, Switz.: Springer
33. Pan J, Yan Q. 2018. *J. Semicond.* 39:071001
34. Holby EF, Wang G, Zelenay P. 2020. *ACS Catal.* 10(24):14527–39
35. White A. 2012. *MRS Bull.* 37(8):715–16
36. Jain A, Ong SP, Hautier G, Chen W, Richards WD, et al. 2013. *APL Mater.* 1:011002
37. Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, et al. 2012. *Comput. Mater. Sci.* 58:227–35
38. Saal JE, Kirklin S, Aykol M, Meredig B, Wolverton C. 2013. *JOM* 65(11):1501–9
39. Verduzco Gastelum JC, Strachan A. 2021. *Citrine tools for materials informatics*. Software Package, nanoHub. **https://doi/org/10.21981/DRKB-XX94**
40. Gražulis S, Chateigner D, Downs RT, Yokochi A, Quirós M, et al. 2009. *J. Appl. Crystallogr.* 42(4):726–29
41. Winther KT, Hoffmann MJ, Boes JR, Mamun O, Bajdich M, Bligaard T. 2019. *Sci. Data* 6:75
42. Blaiszik B, Chard K, Pruyne J, Ananthakrishnan R, Tuecke S, Foster I. 2016. *JOM* 68(8):2045–52
43. Kim S, Chen J, Cheng T, Gindulyte A, He J, et al. 2019. *Nucleic Acids Res.* 47(D1):D1102–D1109
44. Luo J, Zhang S, Sun M, Yang L, Luo S, Crittenden JC. 2019. *ACS Nano* 13(9):9811–40
45. EPA (Environ. Prot. Agency). 2022. *Overview of greenhouse gases*. Environ. Top., EPA, Washington, DC. **https://www.epa.gov/ghgemissions/overview-greenhouse-gases**
46. Benson EE, Kubiak CP, Sathrum AJ, Smieja JM. 2009. *Chem. Soc. Rev.* 38:89–99
47. Jin H, Zhang H, Li J, Wang T, Wan L, et al. 2019. *J. Phys. Chem. Lett.* 10(17):5211–18
48. Barrett J. 2003. *Inorganic Chemistry in Aqueous Solution*. Cambridge, UK: R. Soc. Chem.
49. Xiang C, Weber AZ, Ardo S, Berger A, Chen Y, et al. 2016. *Angew. Chem. Int. Ed.* 55(42):12974–88
50. Ringe S, Hormann NG, Oberhofer H, Reuter K. 2022. *Chem. Rev.* 122(12):10777–820. **https://pubs.acs.org/doi/abs/10.1021/acs.chemrev.1c00675**
51. Schwarz K, Sundararaman R. 2020. *Surf. Sci. Rep.* 75(2):100492
52. Singh AK, Zhou L, Shinde A, Suram SK, Montoya JH, et al. 2017. *Chem. Mater.* 29(23):10159–67
53. Pourbaix M. 1974. *Atlas of Electrochemical Equilibria in Aqueous Solutions*. Houston, TX: Natl. Assoc. Corros. Eng.

54. Back S, Na J, Ulissi ZW. 2021. *ACS Catal.* 11(5):2483–91
55. Karmodak N, Andreussi O. 2020. *ACS Energy Lett.* 5(3):885–91
56. Zhou L, Shinde A, Montoya JH, Singh A, Gul S, et al. 2018. *ACS Catal.* 8(12):10938–48
57. Gunasooriya GTKK, Nørskov JK. 2020. *ACS Energy Lett.* 5(12):3778–87
58. Wang Z, Zheng YR, Chorkendorff I, Nørskov JK. 2020. *ACS Energy Lett.* 5(9):2905–8
59. Persson KA, Waldwick B, Lazic P, Ceder G. 2012. *Phys. Rev. B* 85:235438
60. Koyama M, Zhang Z, Wang M, Ponge D, Raabe D, et al. 2017. *Science* 355(6329):1055–57
61. Kazemi F, Saberi A, Malek-Ahmadi S, Sohrabi S, Rezaie H, Tahriri M. 2011. *Ceram. Silik.* 55:26–30
62. Luo G, Yang S, Jenness GR, Song Z, Kuech TF, Morgan D. 2016. *NPG Asia Mater.* 9:e345
63. Belsky A, Hellenbrandt M, Karen VL, Luksch P. 2002. *Acta Crystallogr. B* 58(3):364–69
64. Sun W, Dacek ST, Ong SP, Hautier G, Jain A, et al. 2016. *Sci. Adv.* 2(11):e1600225
65. Kumari S, Gutkowski R, Junqueira JRC, Kostka A, Hengge K, et al. 2018. *ACS Comb. Sci.* 20(9):544–53
66. Wygant BR, Kawashima K, Mullins CB. 2018. *ACS Energy Lett.* 3(12):2956–66
67. Hubert MA, Patel AM, Gallo A, Liu Y, Valle E, et al. 2020. *ACS Catal.* 10(20):12182–96
68. Stevens MB, Kreider ME, Patel AM, Wang Z, Liu Y, et al. 2020. *ACS Appl. Energy Mater.* 3(12):12433–46
69. Mai H, Le TC, Hisatomi T, Chen D, Domen K, et al. 2021. *iScience* 24(9):103068
70. Kumar R, Singh AK. 2021. *npj Comput. Mater.* 7:197
71. Tao Q, Lu T, Sheng Y, Li L, Lu W, Li M. 2021. *J. Energy Chem.* 60:351–59
72. Agarwal A, Goverapet Srinivasan S, Rai B. 2021. *Front. Mater.* 8:292
73. Mazheika A, Wang YG, Valero R, Viñes F, Illas F, et al. 2022. *Nat. Commun.* 13:419
74. Li X, Maffettone PM, Che Y, Liu T, Chen L, Cooper AI. 2021. *Chem. Sci.* 12(32):10742–54
75. Ding J, Bao J, Sun S, Luo Z, Gao C. 2009. *J. Comb. Chem.* 11(4):523–26
76. Seyler M, Stoewe K, Maier WF. 2007. *Appl. Catal. B Environ.* 76(1–2):146–57
77. Lettmann C, Hinrichs H, Maier WF. 2001. *Angew. Chem. Int. Ed.* 40(17):3160–64
78. Parr RG. 1983. *Annu. Rev. Phys. Chem.* 34:631–56
79. Chase MW Jr. 1998. *NIST-JANAF Themochemical Tables*. J. Phys. Chem. Ref. Data Monogr. 9. Woodbury, NY: Am. Chem. Soc./AIP. 4th ed.
80. Kulik HJ. 2015. *J. Chem. Phys.* 142:240901
81. Garza AJ, Scuseria GE. 2016. *J. Phys. Chem. Lett.* 7(20):4165–70
82. Gerber IC, Angyán JG. 2005. *Chem. Phys. Lett.* 415(1–3):100–5
83. Onida G, Reining L, Rubio A. 2002. *Rev. Mod. Phys.* 74(2):601–59
84. Wu Y, Lazic P, Hautier G, Persson K, Ceder G. 2013. *Energy Environ. Sci.* 6:157–68
85. Castelli IE, Hüser F, Pandey M, Li H, Thygesen KS, et al. 2015. *Adv. Energy Mater.* 5(2):1400915
86. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, et al. 2011. *J. Mach. Learn. Res.* 12:2825–30
87. Bonaccorso G. 2017. *Machine Learning Algorithms: A Reference Guide to Popular Algorithms for Data Science and Machine Learning*. Birmingham, UK: Packt
88. Masood H, Toe CY, Teoh WY, Sethu V, Amal R. 2019. *ACS Catal.* 9(12):11774–87
89. Zhang Q, Chang D, Zhai X, Lu W. 2018. *Chemom. Intel. Lab. Syst.* 177:26–34
90. Biswas T, Singh AK. 2021. *npj Comput. Mater.* 7:189
91. Mounet N, Gibertini M, Schwaller P, Campi D, Merkys A, et al. 2018. *Nat. Nanotechnol.* 13(3):246–52
92. Back S, Tran K, Ulissi ZW. 2020. *ACS Appl. Mater. Interfaces* 12(34):38256–65
93. McCafferty E. 2010. *Introduction to Corrosion Science*. New York: Springer
94. Hansen HA, Rossmeisl J, Nørskov JK. 2008. *Phys. Chem. Chem. Phys.* 10(25):3722–30
95. Mathew K, Singh AK, Gabriel JJ, Choudhary K, Sinnott SB, et al. 2016. *Comput. Mater. Sci.* 122:183–90
96. Boland TM, Singh AK. 2022. *Comput. Mater. Sci.* 207:111238
97. Ong SP, Richards WD, Jain A, Hautier G, Kocher M, et al. 2013. *Comput. Mater. Sci.* 68:314–19
98. Wang Z, Guo X, Montoya J, Nørskov JK. 2020. *npj Comput. Mater.* 6:160
99. Huang LF, Rondinelli JM. 2019. *npj Mater. Degrad.* 3:14
100. Ayodele BV, Alsaffar MA, Mustapa SI, Cheng CK, Witoon T. 2021. *Process Saf. Environ. Prot.* 145:120–32

Annual Review of Condensed Matter Physics

Volume 14, 2023

Contents

Errata

An online log of corrections to *Annual Review of Condensed Matter Physics* articles may be found at http://www.annualreviews.org/errata/conmatphys