ACS Catalysis

pubs.acs.org/acscatalysis

Research Article

Learning Design Rules for Selective Oxidation Catalysts from High-Throughput Experimentation and Artificial Intelligence

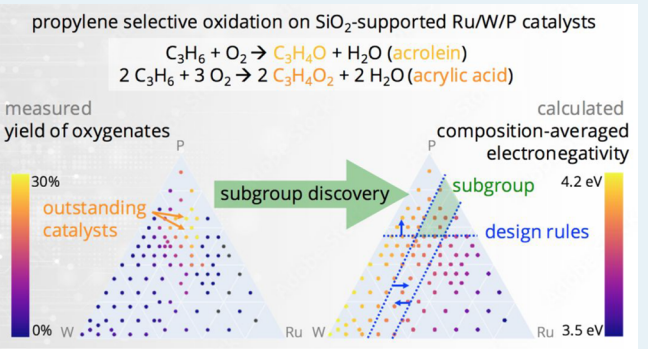
Lucas Foppa,* Christopher Sutton, Luca M. Ghiringhelli, Sandip De,* Patricia Löser, Stephan A. Schunk, Ansgar Schäfer, and Matthias Scheffler

Cite This: *ACS Catal.* 2022, 12, 2223–2232

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The design of heterogeneous catalysts is challenged by the complexity of materials and processes that govern reactivity and by the fact that the number of good catalysts is very small in comparison to the number of possible materials. Here, we show how the subgroup-discovery (SGD) artificial-intelligence approach can be applied to an experimental plus theoretical data set to identify constraints on key physicochemical parameters, the so-called SG *rules*, which exclusively describe materials and reaction conditions with outstanding catalytic performance. By using high-throughput experimentation, 120 SiO_2-supported catalysts containing ruthenium, tungsten, and phosphorus were synthesized and tested in the catalytic oxidation of propylene. As candidate descriptive parameters, the temperature and 10 parameters related to the composition and chemical nature of the catalyst materials, derived from calculated free-atom properties, were offered. The temperature, the phosphorus content, and the composition-weighted electronegativity are identified as key parameters describing high yields toward the value-added oxygenate products acrolein and acrylic acid. The SG rules not only reflect the underlying processes particularly associated with high performance but also guide the design of more complex catalysts containing up to five elements in their composition.



KEYWORDS: *artificial intelligence, subgroup discovery, high-throughput experimentation, selective oxidation, propylene, ruthenium*

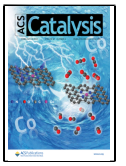
INTRODUCTION

Heterogeneous catalysis is governed by an intricate interplay of multiple processes[1] such as the surface reaction networks and the typically unknown dynamic restructuring of the catalyst material under the reaction conditions. Thus, the design of new materials is challenging. While theoretical approaches attempt to address the complexity of heterogeneous catalysis,[2] the explicit atomistic modeling of the full catalytic progression by first-principles methods is impractical. Another approach for identifying novel catalysts consists of the use of high-throughput experimentation (HTE) to test large numbers of materials.[3] However, utilizing the information obtained by the experiments to decide on the next promising materials to investigate is not straightforward.[4] As the number of possible materials is practically infinite and the number of good catalysts is very small, the direct approach is unlikely to identify the needed catalyst material.

First, when large libraries of materials are tested, the detailed characterization of each material is typically not feasible. Thus, only a small amount of information on the structure and physicochemical properties of each compound might be available. This hinders an in-depth understanding of the underlying processes governing reactivity, which could be used for rational catalyst design. Second, distinct catalytic mechanisms might operate, depending on the materials and reaction conditions, and only very few situations result in good catalytic performance. This leads to an unbalanced distribution between high- and low-performance scenarios and brings into question the usefulness of *global* models to help deciding on the next materials to be tested. These models are trained to describe all materials and reaction conditions simultaneously by minimizing the expected average prediction error over all samples. While this approach may provide an accurate prediction on average, it does not necessarily allow for a focused modeling of the most interesting materials and mechanisms. Alternative approaches for catalyst design are therefore required.

Several recent studies have described artificial-intelligence approaches based on physicochemical parameters for the

Received: October 18, 2021
Revised: January 7, 2022
Published: January 31, 2022



https://doi.org/10.1021/acscatal.1c04793
ACS Catal. 2022, 12, 2223–2232

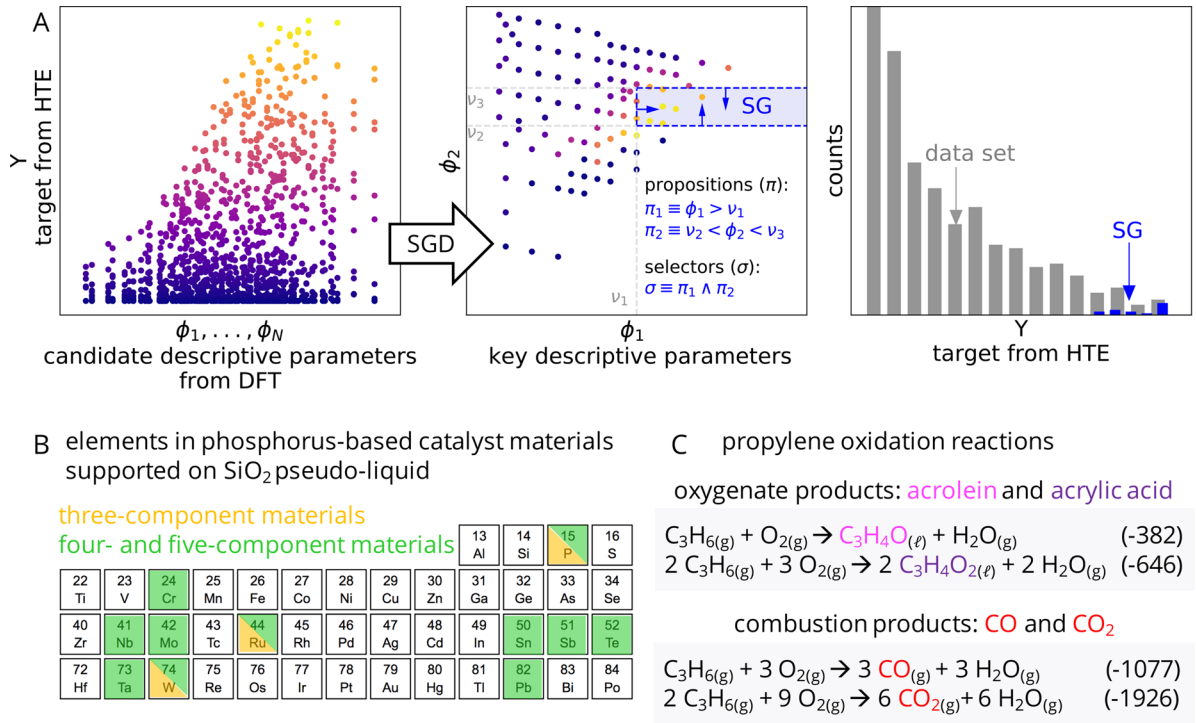


Figure 1. (A) SGD approach for identifying key descriptive parameters and rules associated with materials and reaction conditions with outstanding catalytic performance. The rules are given by the propositions and consist of constraints on the values of key descriptive parameters. "∧" denotes the "AND" connector. (B) Elements entering the composition of the SiO_2-supported materials. (C) Competing reactions in propylene oxidation leading not only to the desired oxygenates but also to the combustion byproducts. The values shown in parentheses in (C) are the reaction enthalpies, in kJ/mol.[9]

analysis and discovery of catalytic systems.[5] Here, we apply the subgroup-discovery (SGD) artificial-intelligence local approach[6] to a hybrid data set obtained from HTE and theory to identify key physicochemical descriptive parameters and constraints on their values, i.e., rules, which are particularly associated with high performance. The reactivity measured by HTE is used as a target in the SGD analysis. The temperature- and composition-dependent physicochemical properties evaluated with density functional theory (DFT) calculations are used as candidate descriptive parameters.

The SGD approach has been applied in computational catalysis[7] and materials science.[6e,8] It starts with the generation of a pool of propositions (π), statements about the data that apply only to a portion of the data set. For the case of continuous candidate descriptive parameters, the propositions are inequalities constraining their values. Then, SGD identifies selectors (σ), i.e., statements formed by a number of propositions and the "AND" connector (denoted "∧"), that result in the selection of subgroups of materials and conditions with the most outstanding distributions of the target values with respect to the whole data set (Figure 1A). The propositions entering these selectors can be seen as rules describing the exceptional SG behavior. The parameters entering these propositions are in turn the key, most relevant descriptive parameters, out of all the offered parameters, associated with the desired reactivity. Because the SG search is performed by maximizing a quality function that measures how outstanding specific subselections of data points are, this approach identifies a local behavior. Thus, the identified rules reflect the specific underlying processes resulting in outstanding performance.

We apply the SGD-HTE and theory approach to the selective oxidation of propylene on SiO_2-supported catalysts based on ruthenium, tungsten, and phosphorus. By using the product yield measured by HTE as a target, we circumvent the need for the explicit modeling of the full catalytic progression. Additionally, because the candidate descriptive material parameters can be calculated by first-principles methods, extensive material characterization by experiment is not required and the resulting SG rules can be used to identify promising catalyst candidates which have not yet been synthesized by experiment.

SELECTIVE OXIDATION REACTION AND HIGH-THROUGHPUT EXPERIMENTATION

The selective partial oxidation of light alkanes to value-added olefins or oxygenates is an efficient route for feedstock upgrading.[10] However, the intricate surface reaction networks[11] typically lead to product mixtures containing up to 20 different molecules, including undesirable byproducts such as CO and CO_2. In order to selectively produce the olefins or the oxygenates, mixed-metal oxide or phosphate heterogeneous catalysts based on molybdenum and vanadium redox-active species have been used, such as $MoVTeNbO_x$ and the state of the art industrial catalyst for *n*-butane selective oxidation, vanadyl pyrophosphate. Several recent investigations have explored the physicochemical properties and the catalytic activity of mixed-metal phosphates in a systematic way.[12]

Platinum-group-metal-based catalysts commonly result in hydrocarbon combustion products. However, ruthenium-based materials also catalyze the partial oxidation of methane to CO.[13] Moreover, the isolation of ruthenium species was proposed as a strategy to increase the catalyst selectivity in

https://doi.org/10.1021/acscatal.1c04793
ACS Catal. 2022, 12, 2223−2232

oxidation reactions.[14] Analogously, it was shown that the isolation of vanadium species in a tungsten phosphate matrix increases the catalyst selectivity toward oxygenates in the *n*-butane oxidation reaction.[15] In this study, we investigate materials based on ruthenium combined with tungsten and phosphorus (Figure 1B) as an alternative class of catalysts for selective oxidation. The combination of Ru with tungsten and phosphorus, in a tungsten phosphate like matrix, could favor selectivity toward the desired olefins and oxygenates, following a catalyst design strategy based on the dilution of highly active metal sites. With the aim of studying these systems, HTE measurements were performed using 120 different *three-component* catalyst compositions containing ruthenium, tungsten, and phosphorus in different proportions. At each catalyst composition, several reaction temperatures between 200 and 400 °C were examined. The detailed preparation, characterization, and reactivity analysis of these catalysts in the selective oxidation of *n*-butane, propane, and propylene is discussed in a separate contribution.[16] In this paper, we only provide details of the propylene selective oxidation reaction (Figure 1C).

All of the reactions were carried out in tubular, fixed-bed reactors with the following reaction feed: Ar, H_2O, N_2, O_2, and propylene (C_3H_6) with molar rates of 4.015, 4.015, 104.40, 20.08, and 1.57 mmol/h, respectively. The same mass of catalyst was used in all reactions, so that the contact time, in terms of volumetric flow per mass of catalyst, was kept fixed across experiments. These three-component catalysts were prepared on a SiO_2 pseudoliquid support and might present a disordered, possibly amorphous, structure. The atomic structures of all the tested catalysts are not known in detail. However, similar catalytic performance was found for crystalline and disordered phases at the same composition.[16] This indicates that the composition is more crucial for the catalytic performance than the degree of crystallinity.

In HTE, a large materials space is accessible for catalyst design by changing the relative amount of each component and the specific elements on the catalyst composition. Approaches to guide the efficient exploration of such a materials space, indicating the most promising compositions to be tested next, are thus desirable. The most interesting compositions are those that display both considerable activity, i.e., those providing significant propylene conversion, and selectivity, i.e., those that specifically form the desired oxygenates (acrolein and acrylic acid, Figure 1C) from propylene. This is motivated by using the yield of oxygenates $Y_{\text{oxygenates}}$ as target in our SGD analysis, defined as

$$Y_{\text{oxygenates}} = Y_{\text{acrolein}} + Y_{\text{acrylic acid}} = \frac{\dot{F}_{\text{acrolein,out}}}{\dot{F}_{\text{propylene,in}}} + \frac{\dot{F}_{\text{acrylic acid,out}}}{\dot{F}_{\text{propylene,in}}} \quad (1)$$

In eq 1, $\dot{F}_{\text{A,in}}$ and $\dot{F}_{\text{A,out}}$ denote the molar rate, in mmol/h, of species A in the reactor feed and outlet, respectively. Our goal is to identify key parameters and rules describing materials and reaction conditions that give high yields of oxygenates.

SUBGROUP DISCOVERY APPROACH

The two main crucial aspects in SGD are the offered candidate descriptive parameters and the quality function. In this work, we use the reaction temperature (T) and the phosphorus molar content (x_p) as experimental candidate parameters. In addition, we include a set of free-atom properties as candidate descriptive parameters to characterize the catalyst material in terms of the proportion and chemical nature of the elements entering the composition. The following elemental properties are used:

- the radii of maximum electron density of *s*, *p*, *d*, and valence orbitals (r_s, r_p, r_d, and r_{val}, respectively)
- the Kohn−Sham single-particle eigenvalues of the highest occupied and lowest unoccupied states (ε_{H} and ε_{L}, respectively)
- the electron affinity (EA)
- the ionization potential (IP)
- the electronegativity (EN), defined as $\text{EN} = \frac{\text{EA} + \text{IP}}{2}$.

These properties were calculated for the isolated atoms using DFT-PBEsol[17] and the FHI-aims[18] code (further calculation details and values for the elemental properties used in the work are available in the Supporting Information). r_{val} is defined as the radius of the highest occupied state. For a given catalyst composition, the per-element free-atom properties are converted into system-specific properties by taking the composition-weighted average

$$\overline{\varphi_a} = \sum^{M} \varphi_{a,i} x_i \quad (2)$$

where φ_a is an arbitrary elemental property, x_i is the molar content of element i in the material, and i runs over all M elements in the composition. For the three-component materials, $M = 3$. We note that oxygen is also present in all materials, but its proportion is not known from the catalyst formulation nor measured for all materials. Therefore, the oxygen content is not included in the material's characterization. Properties that can be readily calculated for the free atom are advantageous to structure-based properties because they do not have to be re-evaluated for each new material. Furthermore, it should be noted that the composition-weighted average is defined for an arbitrary number of components (or elements). Therefore, the key descriptive parameters identified by the SGD of three-component materials can be used to design materials containing different elements or more than three components (*vide infra*). This would not be the case if only composition parameters (e.g., x_{P}, x_{Ru}, and x_{W}) were used, since these quantities are not defined for materials containing elements different from Ru, P, and W or more than three components. In total, 11 descriptive parameters are used in our SGD analysis: T, x_{P}, $\overline{r_s}$, $\overline{r_p}$, $\overline{r_d}$, $\overline{r_{\text{val}}}$, $\overline{\varepsilon_{\text{H}}}$, $\overline{\varepsilon_{\text{L}}}$, $\overline{\text{EA}}$, $\overline{\text{IP}}$, and $\overline{\text{EN}}$.

As the SGD quality function, we use

$$Q(\text{P, SG}) = \frac{s(\text{SG})}{s(\text{P})} D_{\text{cJS}}(\text{P, SG}) \quad (3)$$

where the *coverage* $\frac{s(\text{SG})}{s(\text{P})}$ is the ratio between the number of data points in the subgroup, s(SG), and the total number of data points in the whole data set, s(P), and D_{cJS}(P,SG) is the cumulative Jensen−Shannon divergence between the distribution of the target values in the SG and the distribution of the target values in the whole data set.[19] The coverage term controls the subgroup size and prevents very small SGs with little statistical significance from being selected. The second term, D_{cJS}, is the cumulative-distribution-function formulation[19] of the Jensen−Shannon divergence, which is a properly symmetrized version of the information-theoretical Kullback−

https://doi.org/10.1021/acscatal.1c04793
ACS Catal. 2022, 12, 2223−2232

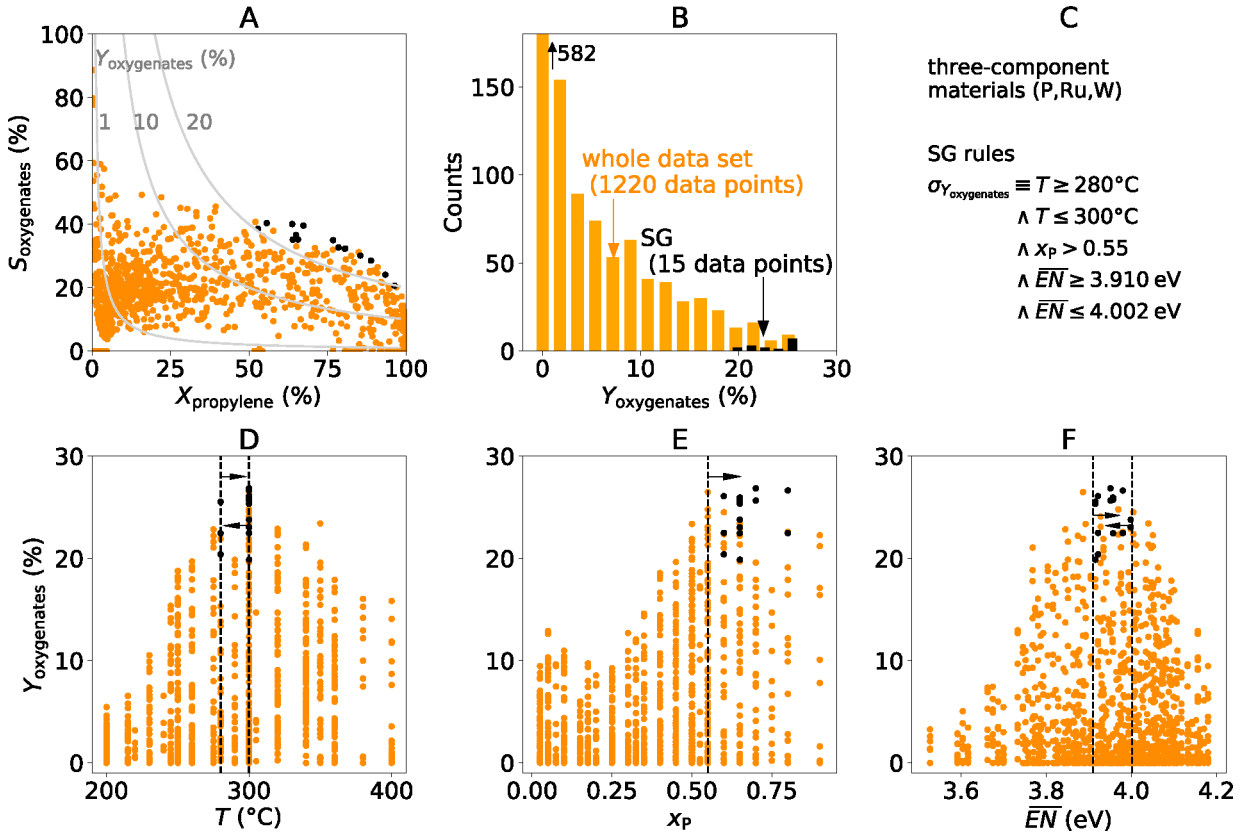


Figure 2. SGD analysis of propylene selective oxidation using three-component materials with ruthenium, tungsten, and phosphorus: (A) overview of reactivity measured by HTE; (B) distribution of oxygenate yield over the data set of 1220 measurements; (C) identified rules describing the SG; (D−F) SG rules (indicated by the black dashed lines and arrows) on the identified key descriptive parameters temperature (T), phosphorus molar content (x_P), and composition-averaged electronegativity ($\overline{EN}$), respectively. The data points corresponding to the identified SG are displayed in black. The propylene conversion and the selectivity toward oxygenates are defined by $X_{propylene} = \frac{\dot{F}_{propylene,in} - \dot{F}_{propylene,out}}{\dot{F}_{propylene,in}}$ and $S_{oxygenates} = \frac{\dot{F}_{acrolein,out} + \dot{F}_{acrylic\ acid,out}}{\dot{F}_{propylene,in} - \dot{F}_{propylene,out}}$, respectively.

Leibler divergence. D_{cJS} measures the dissimilarity between two distributions: it assumes close to zero values for similar distributions and increases, for instance, as the distributions have different standard deviations or different means. Thus, the second term in eq 3 favors the identification of SGs presenting target values as "unusual" as possible in comparison to the majority of the observations. It also favors distributions that are contained in narrower value ranges in comparison to the whole data set. When most of the data points at hand contain low-performing materials and conditions, the use of D_{cJS} in the quality function allows focusing on the exceptionally high performing materials.

The SGD approach contrasts with conventional artificial-intelligence methods such as decision trees, which are based on the optimization of a function that measures the global performance of the model across the whole data set (e.g., mean absolute error or root mean squared error). While global approaches may provide a good description on average, they do not focus on the outstanding data points. We also note that SGD identifies the key, most relevant descriptive parameters out of many candidate descriptive parameters. Conversely, conventional regression approaches (e.g., multivariate regression) exploit all of the parameters simultaneously. As a consequence, the most important parameters are not necessarily determined by such an analysis. This is problematic when obtaining the candidate descriptive parameters for new materials (extrapolation) involves significant experimental or computational effort. Further SGD details, including a detailed description of the approach and of the Jensen−Shannon divergence are available in Supporting Information. SGD was compared to the decision-tree approach in previous works by some of us.[7]

SUBGROUP OF THREE-COMPONENT CATALYSTS WITH EXCEPTIONAL PERFORMANCE

The propylene conversion vs oxygenates selectivity profiles and the distribution of yield of oxygenates in the data set (Figure 2A,B, respectively) show that the vast majority of observations correspond to low performance. Indeed, 50% of the measured materials and conditions result in a less than 2% oxygenate yield and only 41 measurements, out of 1220, are associated with yields of oxygenates above 20%. The average oxygenate yield over the whole data set is equal to 4.83%, and the maximum $Y_{oxygenates}$ value is 26.85%.

By applying the SGD, we identified several SGs providing near-optimal quality-function values (Figure S3). Among the SGs displaying quality-function values within 40% of the optimal value, we selected, for further analysis and discussion, the SG that presents the highest value of cumulative Jensen−Shannon divergence (0.693). This SG contains only 15 data points, i.e., approximately 1.2% of the data set, which all have high yields of oxygenates (Figure 2A,B, in black). The average

https://doi.org/10.1021/acscatal.1c04793
ACS Catal. 2022, 12, 2223−2232

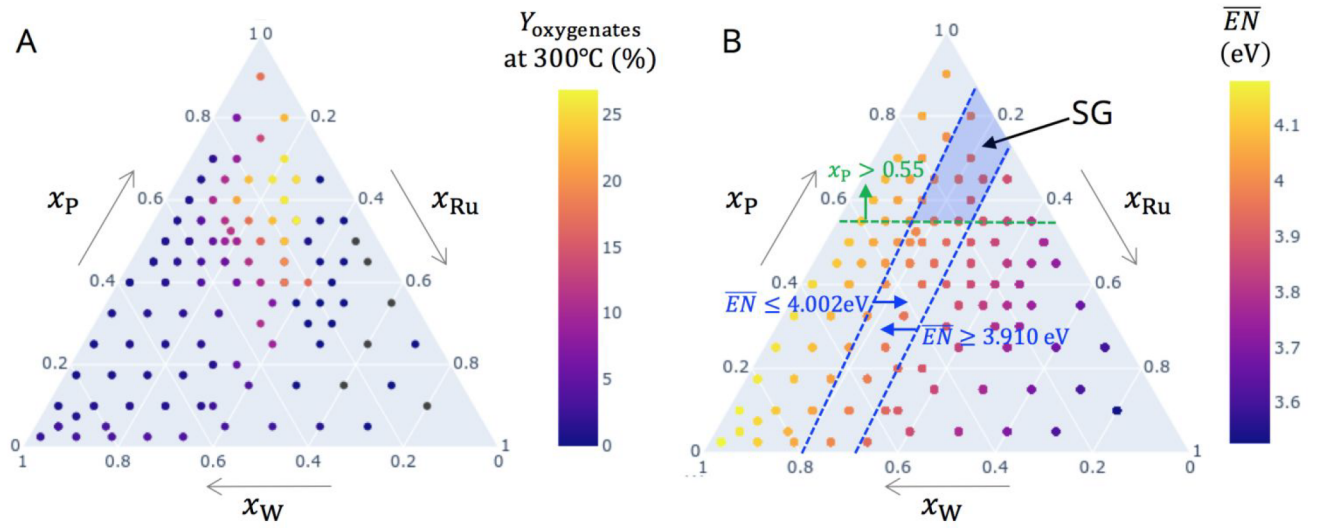


Figure 3. Ternary diagrams for three-component materials with ruthenium, tungsten, and phosphorus tested in propylene selective oxidation using HTE: (A) measured yield of oxygenates at 300 °C; (B) composition-averaged electronegativity $\overline{EN}$ for each tested composition. The SG rules are shown by the dashed lines and arrows in (B), and the portion of the ternary diagram selected by the SG rules is shown in blue.

yield of oxygenates in this SG is equal to 24.15%: i.e., 5 times higher than the average on the whole data set. This SG is described by rules on three descriptive parameters: $280 \leq T \leq 300$ °C, $x_P > 0.55$, and $3.910 \leq \overline{EN} < 4.002$ eV (Figure 2C).

The rule on the temperature highlights that the highest yields of oxygenates are observed for intermediate temperatures within the tested range of 200−400 °C (Figure 2C). This could be related to the fact that the yield of oxygenates is favored at intermediate propylene conversions (Figure S4). Indeed, the temperature is the descriptive parameter that correlates the most with the propylene conversion, of all the candidate parameters that were offered (Figure S2). The rule on the phosphorus content shows that a relatively high phosphorus content is needed to achieve outstanding performance (Figure 2D). This could be related to the dilution of ruthenium sites on a phosphate matrix that occurs at high phosphorus loadings.[15,16] Finally, the rule on the composition-averaged electronegativity (Figure 2F) effectively limits the range of Ru contents, as shown in the ternary diagram of Figure 3B. This reflects the fact that Ru is needed to achieve propylene conversion (Figure S5A) but that too much of this element in the composition leads to undesired combustion products (Figure S5B). The electronegativity of an element reflects its tendency to attract electronic density in a chemical bond. Thus, from a physicochemical standpoint, the relevance of $\overline{EN}$ could be related to the strength and nature of certain bonds within the materials or between the materials' surfaces and reacting species: for instance, metal−oxygen bonds.[20] These bonds are crucial in several processes taking place during the oxidation reaction, such as such as the O_2 dissociation and the oxygen transfers from the catalyst surface to adsorbed organic species in order to form the C−O bonds of acrolein and acrylic acid. However, we note that $\overline{EN}$ is an effective (mean field) electronegativity and not a specific electronegativity of a certain element. Thus, $\overline{EN}$ is also related to the composition.

We wish to stress that the rules derived by SGD depend on the *combination of constraints* on the parameters T, x_P, and $\overline{EN}$ as a descriptor for outstanding performance. Therefore, by assigning a too specific, chemical meaning to each parameter individually, one might overlook the possibly intricate interplay of the many processes governing catalysis. Moreover, the interpretation presented above is speculative in the sense that it is based on the current knowledge about the catalyst materials and reaction. It is possible, or even likely, that other, so far not well understood or unknown underlying processes (e.g., the dynamic restructuring of the catalyst during the reaction[21]) play a significant role in determining the outstanding behavior. The SG rules might capture these processes. In fact, the SG rules do not necessarily reflect causality. Thus, the physical relationship between the identified parameters and the underlying chemistry might be indirect.

Similar SG rules are obtained when the training is performed with randomly selected 90% of the data points (see cross-validation study in the Supporting Information) or when the data points presenting yield of oxygenates lower than 3% are excluded from training (see details in the Supporting Information). SG rules constraining the $\overline{EN}$ parameter to an intermediate range, for instance, are always observed when only 90% of the data is used for training. Furthermore, the ranges of variation of minimum and maximum thresholds are [3.882, 3.910 eV] and [3.989, 4.031 eV], respectively: i.e., similar to the thresholds shown in Figure 2F. These results indicate that the SG rules are not strongly affected by variations of the data used for their derivation. We have also verified that the SG rules derived on the basis of 90% of the data set (training sets) are able to select the outstanding materials in the remaining 10% (test sets) (see the Supporting Information for details).

For comparison, we have also performed the SGD using only the experimental candidate descriptive parameters T, x_P, and x_{Ru}. We identify the SG rules $300 \leq T \leq 300$ °C, $x_P > 0.55$, and $0.15 \leq x_{Ru} \leq 0.25$. These SG rules select 13 data points that correspond to a cumulative Jensen−Shannon divergence of 0.704. The quality-function value associated with the selected data points and some of the SG rules are similar to those discussed in Figure 2. However, the rules derived solely on the basis of experimental descriptive parameters are specific to three-component materials composed of phosphorus, ruthenium, and tungsten. Thus, it is not straightforward to

https://doi.org/10.1021/acscatal.1c04793
ACS Catal. 2022, 12, 2223−2232

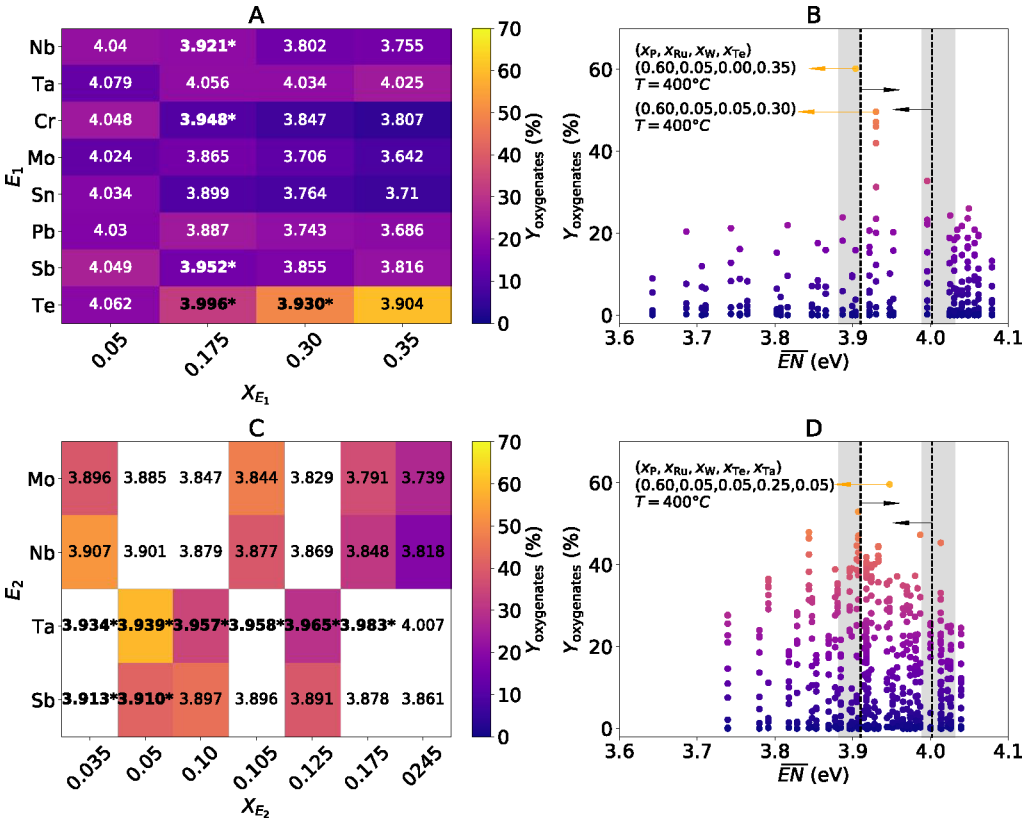


Figure 4. SG rules applied to the design of four- and five-component materials for propylene selective oxidation: (A, C) composition-averaged electronegativity ($\overline{EN}$) for different elements and molar contents in four- and five-component materials, respectively; (B, D) distribution of all measured yields of oxygenates (214 and 533 data points) for four- and five-component materials, respectively. In (A) and (C), the $\overline{EN}$ values are shown in boldface and are marked with asterisks if they satisfy the SG rules on $\overline{EN}$ identified on the basis of the three-component materials. The colors in (A) and (C) indicate the highest measured yield of oxygenates for each material. The SG rules identified on the basis of the three-component materials are indicated by the dashed lines and arrows in (B) and (D). The shaded areas in (B) and (D) indicate the variability of $\overline{EN}$ thresholds observed when different subsets of the data, containing only 90% of the data set, are used for training (Table S2). The oxygenate yields shown in (C) correspond to materials with $x_W = 0.035$ for the cases E_2 = Mo, Nb and with $x_W = 0.050$ for the cases E_2 = Ta, Sb. The white cells in (C) indicate materials not measured by HTE.

use these rules for the design of materials containing other elements or more than three components. Conversely, the composition-weighted parameters derived by electronic-structure calculations are well-defined for materials containing an arbitrary number of elements, including elements that are not initially present in the data set. Thus, the rules associated with the theoretical parameters can be exploited for the design of more complex materials (*vide infra*).

Overall, our results demonstrate the ability of the HTE and theory SGD approach to detect interpretable, chemically meaningful, and complex patterns associated with very few data points presenting exceptional catalytic performance.

EXPLOITING THE SUBGROUP RULES FOR THE DESIGN OF FOUR- AND FIVE-COMPONENT CATALYSTS

Using the rules defining the SG of outstanding oxygenate production for the three-component data, we designed more complex materials containing additional elements. We start by considering four-component materials containing ruthenium, tungsten, phosphorus, and one additional E_1 element. For this analysis, we fix the phosphorus content to 0.60 according to the rule identified in Figure 2C. To further reduce the number of possible variables determining the catalyst composition, we also fix the ruthenium molar content to 0.05. We focus on such relatively low ruthenium loadings to ensure that the formation of combustion products is not significant. In this way, the compositions of the four-component materials are determined solely by the choice of E_1 element and its molar content. Materials with an E_1 molar content of 0.35, which do not contain tungsten and are thus composed by three elements, are also referred to as four-component materials in our analysis to highlight that they contain different chemical elements in comparison to ruthenium, tungsten, and phosphorus, the elements present in the materials of the data set used to derive the rules.

We concentrate on E_1 elements that show octahedral coordination patterns among the reported phosphorus-containing material structures[22] and that have a maximum atomic radius difference in comparison to tungsten of 0.10 Å (see details in the Supporting Information). This is to ensure that only elements that are compatible with tungsten, i.e., that could possibly replace tungsten in the material structure, are taken into account. The following E_1 elements are considered: niobium, tantalum, chromium, molybdenum, tin, antimony, and tellurium. These elements have atomic radii of 1.45, 1.45, 1.40, 1.45, 1.45, 1.45, and 1.40 Å, respectively. The atomic radius of tungsten is 1.45 Å. We have also included lead in this analysis, since materials containing this element were also experimentally tested (see below).

https://doi.org/10.1021/acscatal.1c04793
ACS Catal. 2022, 12, 2223−2232

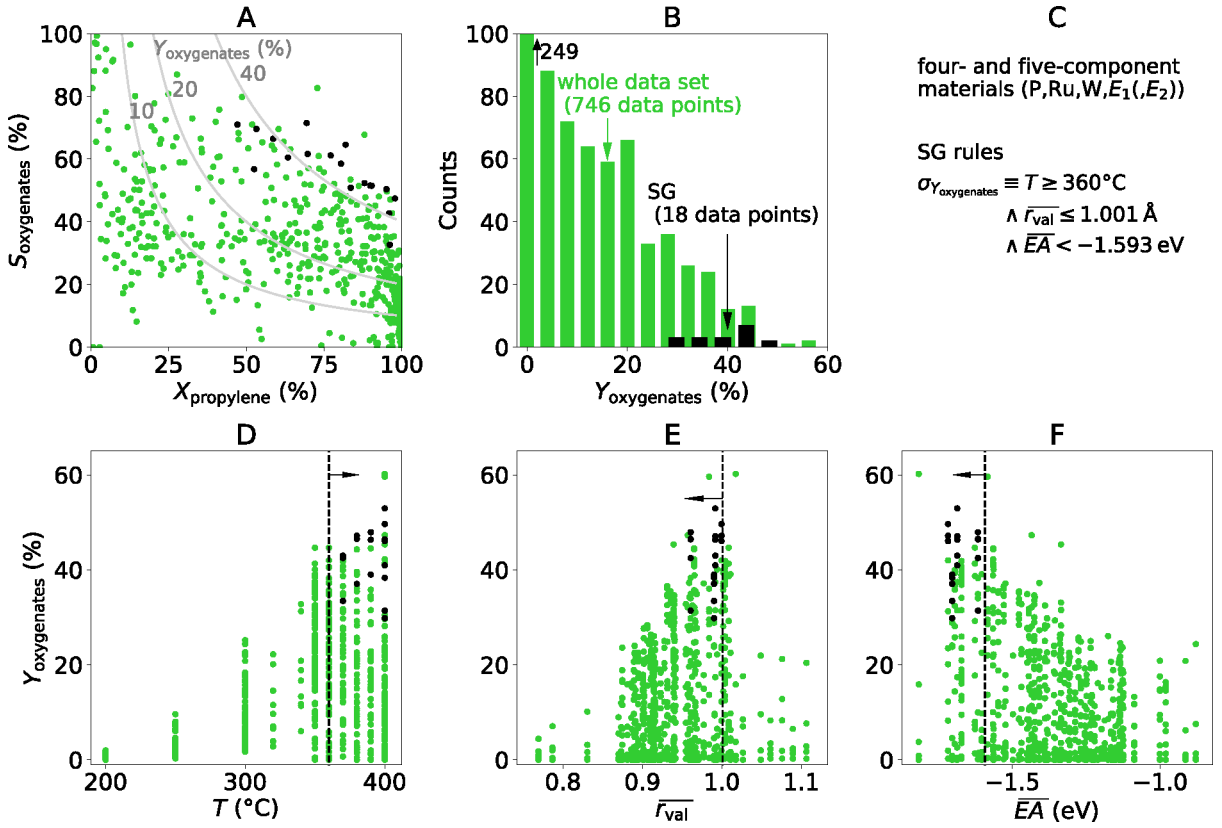


Figure 5. SGD analysis of propylene selective oxidation on four- and five-component materials: (A) overview of reactivity measured by HTE; (B) distribution of oxygenate yield over the data set of 746 measurements; (C) identified rules describing the SG; (D−F) SG rules (indicated by the black dashed lines and arrows) on the identified key descriptive parameter: temperature (T), composition-averaged valence radius ($\overline{r_{val}}$), and electron affinity ($\overline{EA}$), respectively. The data points corresponding to the SG are displayed in black.

We evaluated the composition-averaged electronegativity for the selected E_1 and the molar contents 0.05, 0.175, 0.30, and 0.35 in Figure 4A. In this figure, the $\overline{EN}$ values for the new four-component materials are shown in boldface and marked with asterisks if they satisfy the SG rule $3.910 \leq \overline{EN} < 4.002$ eV. This *catalyst map* suggests that the use of 5.0−17.5 mol % of niobium, chromium, molybdenum, tin, lead, and antimony, in the catalyst composition in addition to ruthenium, tungsten, and phosphorus results in catalysts which are part of the identified SG. Thus, these are likely high performing materials. For the case of tantalum and tellurium, 17.5 mol % or more of these elements is needed for the resulting materials to present $\overline{EN}$ values compatible with the SG.

The four-component catalyst compositions shown in Figure 4A were tested in propylene oxidation using HTE under the same reaction conditions as those used for testing the three-component materials. The highest yield of oxygenates achieved for each composition is shown by the colors in Figure 4A. A comparison of the experimental results with the SG rules on $\overline{EN}$ shows that the catalyst design rules derived by SGD correctly describe the experimental trend. In particular, the materials based on niobium, chromium, molybdenum, tin, lead, and antimony achieve the highest oxygenate yields at relative lower E_1 molar fractions in comparison to the tantalum and tellurium-based materials, in line with the optimal ranges of $\overline{EN}$ values indicated by the SG rules.

All measured yields of oxygenates, corresponding to the materials shown in the catalyst map of Figure 4A tested at several temperatures, are plotted as a function of $\overline{EN}$ in Figure 4B. In this figure, the SG rules on $\overline{EN}$ are shown as vertical black lines and arrows. The variability of $\overline{EN}$ thresholds in the SG rule with respect to the input data set is indicated by the ranges of $\overline{EN}$ values in the gray shaded areas. These ranges correspond to the variations of the thresholds observed when different subsets of data, containing only 90% of the data set, are used for training (see Table S2). The catalyst achieving the highest yield of oxygenates (60.19% at 400 °C) contains a 0.35 molar fraction of tellurium as the E_1 element and lies within the window of $\overline{EN}$ values suggested by the SG rules.

We have also used the SG rules derived from the three-component materials to address five-component materials, which were tested experimentally (Figure 4C,D). For this purpose, E_1 was fixed to be tellurium on the basis of the best four-component catalysts and molybdenum, niobium, tantalum, and antimony were evaluated as E_2. Thus, ruthenium, tungsten, phosphorus, tellurium, and E_2 enter in the composition of the considered five-component materials. The agreement between the SG rule and the measured oxygenate yield is reasonable in spite of the much higher material complexity with respect to the catalysts used for training. In particular, the five-component catalyst corresponding to the highest yield of oxygenates (59.60% at 400 °C) contains tantalum as the E_2 element and the composition-averaged electronegativity for this material is 3.947 eV. Such an $\overline{EN}$ value lies within the threshold defined by the SG rule (Figure 4D).

https://doi.org/10.1021/acscatal.1c04793
ACS Catal. 2022, 12, 2223−2232

These results demonstrate the potential of the SGD-HTE and theory approach to identify generalizable rules describing exceptional performance. Indeed, the identified four- and five-component catalysts are significantly more complex than those of the training data set (three-component materials). Moreover, the outstanding four- and five-component catalysts achieve oxygenate yields (60.19 and 59.60%, respectively) up to twice as large as those obtained with three-component materials (highest value of 26.85%). Therefore, the SG rules hinted at materials that are significantly better performing that any of the materials used in training.

We note that the four- and five-component materials achieve the highest yields of oxygenates at higher temperatures (400 °C) than the three-component systems (300 °C) (see Figure S4). One of the SG rules on the reaction temperature derived on the basis of the three-component materials data set ($T \leq$ 300 °C) is thus not transferable to the four-component set, since this aspect is not included in the training. The temperature is, however, a less crucial parameter than the composition-related parameters within our HTE approach, since the screening of different temperatures is less resource consuming than the screening of different materials, or compositions. For this reason, the four- and five-component materials were tested using the same range of temperatures used for the three-component material (200−400 °C).

Finally, we applied the SGD approach to the four- and five-component HTE data (746 data points, Figure 5A,B) using the same candidate descriptive parameters used for the previous SGD analysis of three-component materials. The identified SG presenting the highest D_{cjS} value (0.355) contains 18 data points: i.e., ca. 2.4% of the data set (black points in Figure 5A,B). The selected data points correspond to one four-component material with tellurium as the E_1 element as well as different compositions of five-component materials with E_1 = Te and E_2 = Mo, Nb. The rules describing this SG (Figure 5C) constrain the values of three parameters: $T \geq 360$ °C, $\overline{r_{val}} \leq 1.001$ Å, and $\overline{EA} < -1.593$ eV, (Figure 5D−F, respectively). The comparison of these SG rules with those for the SG obtained with the three-component materials data set (Figure 2C) highlights the higher temperatures needed for the four- and five-component materials to achieve outstanding performance. Moreover, different composition-dependent parameters ($\overline{r_{val}}$ and $\overline{EA}$) are required to describe this SG in comparison to the case of three-component materials (x_p and $\overline{EN}$), even though the electronegativity and the electron affinity are related by $EN = \frac{EA + IP}{2}$.

We note that two data points presenting high yield of oxygenate are not captured by the SG rules in Figure 5. This could indicate that these two data points are governed by different underlying processes in comparison to the situations belonging to the identified SG. However, these two points fall close to the threshold of the SG rules (Figure 5E,F) and the precise thresholds present some variability with respect to the data used for their derivation (see discussion above on $\overline{EN}$ thresholds).

The SG rules derived in this study are expected to describe outstanding materials whose performance is governed by the same processes governing the reactivity of the exceptional materials in the input data sets used for training. The analysis of four- and five-component materials was focused, nevertheless, on low ruthenium contents and on E_1 and E_2 elements compatible with tungsten: i.e., with similar atomic radii. Thus, it is unclear if the SG rules presented in Figure 5C can identify exceptional materials and conditions for any arbitrary ruthenium content or for E_1 and E_2 elements that have significantly different radii in comparison to tungsten. This is because different mechanisms may operate on these materials that could also lead to exceptional performance. Therefore, the SGD analysis might need to be performed by including new data points covering such thus far unexplored portions of the materials space to enlarge the domain in which the SG rules can detect exceptional catalysts and reaction conditions.

CONCLUSIONS

In this paper, we applied the SGD approach to the design of selective oxidation phosphorus-containing supported catalysts on the basis of data from HTE and DFT calculations. The yield of value-added oxygenate product measured by HTE was used as a target, and parameters obtained from DFT-calculated free-atom properties were offered as candidate descriptive parameters. The composition-weighted electronegativity, the phosphorus content, and the temperature were identified as key parameters associated with an outstanding production of acrolein and acrylic acid from propylene in three-component catalysts containing ruthenium, tungsten, and phosphorus. The SG rules on these key parameters not only rationalize a local reactivity pattern particularly associated with exceptional catalytic performance but also guide the design of more complex catalysts. In particular, a five-component material containing ruthenium, tungsten, phosphorus, tellurium, and tantalum in the composition, which presents an oxygenate yield more than twice as large as any observation in the data set used for training, is captured by the SG rules. This local modeling approach is suitable for the search of exceptional materials whose structures and functions can hardly be modeled explicitly by theory.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.1c04793.

DFT calculation details, additional SGD details, and details on the choice of compatible elements for the four- and five-component materials (PDF)

High-throughput-experimentation data set (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Lucas Foppa − *The NOMAD Laboratory, Fritz-Haber-Institut der Max-Planck-Gesellschaft, D-14195 Berlin, Germany; The NOMAD Laboratory, Humboldt-Universität zu Berlin, D-12489 Berlin, Germany;* orcid.org/0000-0003-3002-062X; Email: foppa@fhi-berlin.mpg.de

Sandip De − *BASF SE, D-67065 Ludwigshafen, Germany*; Email: sandip.de@basf.com

Authors

Christopher Sutton − *The NOMAD Laboratory, Fritz-Haber-Institut der Max-Planck-Gesellschaft, D-14195 Berlin, Germany;* Present Address: Department of Chemistry and Biochemistry. University of South Carolina, Columbia, South Carolina, United States; orcid.org/0000-0003-1206-8080

https://doi.org/10.1021/acscatal.1c04793
ACS Catal. 2022, 12, 2223−2232

Luca M. Ghiringhelli − *The NOMAD Laboratory, Fritz-Haber-Institut der Max-Planck-Gesellschaft, D-14195 Berlin, Germany; FAIRmat, Humboldt-Universität zu Berlin, D-12489 Berlin, Germany*
Patricia Löser − *hte GmbH, D-69123 Heidelberg, Germany*
Stephan A. Schunk − *BASF SE, D-67065 Ludwigshafen, Germany; hte GmbH, D-69123 Heidelberg, Germany*
Ansgar Schäfer − *BASF SE, D-67065 Ludwigshafen, Germany*
Matthias Scheffler − *The NOMAD Laboratory, Fritz-Haber-Institut der Max-Planck-Gesellschaft, D-14195 Berlin, Germany; The NOMAD Laboratory, Humboldt-Universität zu Berlin, D-12489 Berlin, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acscatal.1c04793

Funding

Open access funded by Max Planck Society.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Mario Boley is acknowledged for helpful discussions. We also thank Thomas A. R. Purcell for providing the calculated elemental properties. L.F. acknowledges funding from the NOMAD CoE (European Union's Horizon 2020 research and innovation program under the grant agreement No. 951786). Funding by BASF SE is gratefully acknowledged. We also acknowledge the productive cooperation and fruitful interaction with BASF SE and hte GmbH. This work was supported by the consortia NFDI4Cat (German Research-Data Infrastructure for Catalysis) and FAIRmat (FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids). The SGD analysis described in this publication can be found in a Jupyter notebook at the *NOMAD Artificial-Intelligence Toolkit* (https://nomad-lab.eu/AItutorials/SGD_Propylene_Oxidation_HTE), where it can be repeated and modified directly in a web browser.

■ REFERENCES

(1) (a) Freund, H.-J.; Meijer, G.; Scheffler, M.; Schlögl, R.; Wolf, M. CO Oxidation as a Prototypical Reaction for Heterogeneous Processes. *Angew. Chem., Int. Ed.* **2011**, *50* (43), 10064−10094. (b) Schlögl, R. Heterogeneous Catalysis. *Angew. Chem., Int. Ed.* **2015**, *54* (11), 3465−3520. (c) Foppa, L.; Ghiringhelli, L. M.; Girgsdies, F.; Hashagen, M.; Kube, P.; Hävecker, M.; Carey, S. J.; Tarasov, A.; Kraus, P.; Rosowski, F.; Schlögl, R.; Trunschke, A.; Scheffler, M. Materials genes of heterogeneous catalysis from clean experiments and artificial intelligence. *MRS Bull.* **2021**, 46.

(2) Reuter, K.; Stampf, C.; Scheffler, M. *Ab Initio* Atomistic Thermodynamics and Statistical Mechanics of Surface Properties and Functions. In *Handbook of Materials Modeling: Methods*; Yip, S., Ed.; Springer Netherlands: 2005; pp 149−194.

(3) (a) Hattrick-Simpers, J.; Wen, C.; Lauterbach, J. The materials super highway: integrating high-throughput experimentation into mapping the catalysis materials genome. *Catal. Lett.* **2015**, *145* (1), 290−298. (b) Farrusseng, D. High-throughput heterogeneous catalysis. *Surf. Sci. Rep.* **2008**, *63* (11), 487−513. (c) Tompos, A.; Sanchez-Sanchez, M.; Végvári, L.; Szijjártó, G. P.; Margitfalvi, J. L.; Trunschke, A.; Schlögl, R.; Wanninger, K.; Mestl, G. Combinatorial optimization and synthesis of multiple promoted MoVNbTe catalysts for oxidation of propane to acrylic acid. *Catal. Today* **2021**, *363*, 45−54.

(4) (a) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **2018**, *559* (7714), 377−381. (b) Williams, T.; McCullough, K.; Lauterbach, J. A. Enabling Catalyst Discovery through Machine Learning and High-Throughput Experimentation. *Chem. Mater.* **2020**, *32* (1), 157−165. (c) Langner, S.; Häse, F.; Perea, J. D.; Stubhan, T.; Hauch, J.; Roch, L. M.; Heumueller, T.; Aspuru-Guzik, A.; Brabec, C. J. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems. *Adv. Mater.* **2020**, *32* (14), 1907801.

(5) (a) Suzuki, K.; Toyao, T.; Maeno, Z.; Takakusagi, S.; Shimizu, K.-I.; Takigawa, I. Statistical Analysis and Discovery of Heterogeneous Catalysts Based on Machine Learning from Diverse Published Data. *ChemCatChem.* **2019**, *11* (18), 4537−4547. (b) Mine, S.; Takao, M.; Yamaguchi, T.; Toyao, T.; Maeno, Z.; Hakim Siddiki, S. M. A.; Takakusagi, S.; Shimizu, K.-I.; Takigawa, I. Analysis of Updated Literature Data up to 2019 on the Oxidative Coupling of Methane Using an Extrapolative Machine-Learning Method to Identify Novel Catalysts. *ChemCatChem.* **2021**, *13* (16), 3636−3655. (c) Li, Z.; Ma, X.; Xin, H. Feature engineering of machine-learning chemisorption models for catalyst design. *Catal. Today* **2017**, *280*, 232−238. (d) Li, Z.; Achenie, L. E. K.; Xin, H. An Adaptive Machine Learning Strategy for Accelerating Discovery of Perovskite Electrocatalysts. *ACS Catal.* **2020**, *10* (7), 4377−4384.

(6) (a) Wrobel, S. In *An algorithm for multi-relational discovery of subgroups, European symposium on principles of data mining and knowledge discovery*; Springer: 1997; pp 78−87. (b) Friedman, J. H.; Fisher, N. I. Bump hunting in high-dimensional data. *Statistics and Computing* **1999**, *9* (2), 123−143. (c) Atzmueller, M. Subgroup discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2015**, *5* (1), 35−49. (d) Boley, M.; Goldsmith, B. R.; Ghiringhelli, L. M.; Vreeken, J. Identifying consistent statements about numerical data with dispersion-corrected subgroup discovery. *Data Min. Knowl. Discovery* **2017**, *31* (5), 1391−1418. (e) Goldsmith, B. R.; Boley, M.; Vreeken, J.; Scheffler, M.; Ghiringhelli, L. M. Uncovering structure-property relationships of materials by subgroup discovery. *New J. Phys.* **2017**, *19* (1), 013031.

(7) (a) Foppa, L.; Ghiringhelli, L. M. Identifying Outstanding Transition-Metal-Alloy Heterogeneous Catalysts for the Oxygen Reduction and Evolution Reactions via Subgroup Discovery. *Top. Catal.* **2021**, DOI: 10.1007/s11244-021-01502-4. (b) Mazheika, A.; Wang, Y.; Valero, R.; Viñes, F.; Illas, F.; Ghiringhelli, L. M.; Levchenko, S. V.; Scheffler, M. Artificial-intelligence-driven discovery of catalyst genes with application to CO_2 activation on semiconductor oxides. *Nat. Commun.* **2022**, *13*, 419.

(8) Sutton, C.; Boley, M.; Ghiringhelli, L. M.; Rupp, M.; Vreeken, J.; Scheffler, M. Identifying domains of applicability of machine learning models for materials science. *Nat. Commun.* **2020**, *11* (1), 4428.

(9) *CRC Handbook of Chemistry and Physics*; CRC Press: 2020; Vol. *101*.

(10) (a) Grasselli, R. K.; Burrington, J. D. Oxidation of Low-Molecular-Weight Hydrocarbons. *Handbook of Heterogeneous Catalysis*; Wiley-VCH: 2008; pp 3479−3489 (b) Grasselli, R. K. Fundamental Principles of Selective Heterogeneous Oxidation Catalysis. *Top. Catal.* **2002**, *21* (1), 79−88. (c) Schlögl, R. Selective Oxidation: From a Still Immature Technology to the Roots of Catalysis Science. *Top. Catal.* **2016**, *59* (17), 1461−1476.

(11) (a) Li, X.; Teschner, D.; Streibel, V.; Lunkenbein, T.; Masliuk, L.; Fu, T.; Wang, Y.; Jones, T.; Seitz, F.; Girgsdies, F.; Rosowski, F.; Schlögl, R.; Trunschke, A. How to control selectivity in alkane oxidation? *Chem. Sci.* **2019**, *10* (8), 2429−2443. (b) Kube, P.; Frank, B.; Schlögl, R.; Trunschke, A. Isotope Studies in Oxidation of Propane over Vanadium Oxide. *ChemCatChem.* **2017**, *9* (18), 3446−3455.

(12) (a) Schulz, C.; Roy, S. C.; Wittich, K.; d'Alnoncourt, R. N.; Linke, S.; Strempel, V. E.; Frank, B.; Glaum, R.; Rosowski, F. α_{II}-$(V_{1-x}W_x)OPO_4$ catalysts for the selective oxidation of *n*-butane to maleic anhydride. *Catal. Today* **2019**, *333*, 113−119. (b) Lister, S. E.; Soleilhavoup, A.; Withers, R. L.; Hodgkinson, P.; Evans, J. S. O.

Structures and Phase Transitions in $(MoO_2)_2P_2O_7$. *Inorg. Chem.* **2010**, *49* (5), 2290−2301. (c) Roy, S. C.; Raguš, B.; Assenmacher, W.; Glaum, R. Synthesis and crystal structure of mixed metal(III) tungstenyl(VI) ortho-pyrophosphates. *Solid State Sci.* **2015**, *49*, 18−28.

(13) York, A. P. E.; Xiao, T.; Green, M. L. H. Brief Overview of the Partial Oxidation of Methane to Synthesis Gas. *Top. Catal.* **2003**, *22* (3), 345−358.

(14) Balkus, K. J.; Eissa, M.; Levado, R. Oxidation of alkanes catalyzed by zeolite-encapsulated perfluorinated ruthenium phthalocyanines. *J. Am. Chem. Soc.* **1995**, *117* (43), 10753−10754.

(15) Welker-Nieuwoudt, C.; Rosowski, F.; Goebel, M.; Glaum, R.; Subrata, C. R.; Hautier, G.; Waroquiers, D.; Naumann d'Alnoncourt, R.; Strempel, V. E.; Linke, S. *Wolframphosphate der ReO_3-Strukturfamilie*; BASF SE: 2016.

(16) Machado, R.; Dimitrakopoulou, M.; Girgsdies, F.; Löser, P.; Xie, J.; Wittich, K.; Weber, M.; Geske, M.; Glaum, R.; Karbstein, A.; Rosowski, F.; Titlbach, S.; Skorupska, K.; Tarasov, A.; Schlögl, R.; Schunk, S. A., Platinum group metal phosphates as catalysts for selective C-H activation of lower alkanes. In preparation, 2022.

(17) Csonka, G. I.; Perdew, J. P.; Ruzsinszky, A.; Philipsen, P. H. T.; Lebègue, S.; Paier, J.; Vydrov, O. A.; Ángyán, J. G. Assessing the performance of recent density functionals for bulk solids. *Phys. Rev. B* **2009**, *79* (15), 155107.

(18) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. *Ab initio* molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **2009**, *180* (11), 2175−2196.

(19) Nguyen, H.-V.; Vreeken, J. In *Non-parametric jensen-shannon divergence, Joint European Conference on Machine Learning and Knowledge Discovery in Databases*; Springer: 2015; pp 173−189.

(20) Moltved, K. A.; Kepp, K. P. The Chemical Bond between Transition Metals and Oxygen: Electronegativity, *d*-Orbital Effects, and Oxophilicity as Descriptors of Metal-Oxygen Interactions. *J. Phys. Chem. C* **2019**, *123* (30), 18432−18444.

(21) Trunschke, A.; Noack, J.; Trojanov, S.; Girgsdies, F.; Lunkenbein, T.; Pfeifer, V.; Hävecker, M.; Kube, P.; Sprung, C.; Rosowski, F.; Schlögl, R. The Impact of the Bulk Structure on Surface Dynamics of Complex Mo-V-based Oxide Catalysts. *ACS Catal.* **2017**, *7* (4), 3061−3071.

(22) Waroquiers, D.; Gonze, X.; Rignanese, G.-M.; Welker-Nieuwoudt, C.; Rosowski, F.; Göbel, M.; Schenk, S.; Degelmann, P.; André, R.; Glaum, R.; Hautier, G. Statistical Analysis of Coordination Environments in Oxides. *Chem. Mater.* **2017**, *29* (19), 8346−8360.

https://doi.org/10.1021/acscatal.1c04793
ACS Catal. 2022, 12, 2223−2232