Modelling the Time-Dependent Reactivity of Catalysts by Experiments and Artificial Intelligence

Jonathan M. Mauß,[1] Klara S. Kley,[1] Rohini Khobragade,[1] Nguyen-Khang Tran,[1] Jacopo de Bellis,[1] Ferdi Schüth,[1] Matthias Scheffler,[2] and Lucas Foppa*[2]
[1] *Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany*
[2] *The NOMAD Laboratory at the Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, 14195 Berlin, Germany*

(Dated: January 31, 2025)

The description of heterogeneous catalysis is challenged by the intricacy of numerous multi-scale processes that govern the performance of catalyst materials. The chemical environment of the catalytic process and the kinetics of structural changes create configurations of typically unknown local geometries and chemistry. These may result in significant changes in activity or selectivity within minutes, hours, or longer, during the so-called induction period. Here, we use experimental data together with a focused artificial-intelligence (AI) approach based on subgroup discovery and symbolic regression to model the evolution of the catalyst reactivity with time on stream. We consider palladium-based alloys synthesized mechanochemically and applied in the selective hydrogenation of concentrated acetylene streams resulting from a hypothetical electric plasma-assisted methane-to-ethylene process. Our AI approach starts with the identification of descriptions of materials and reaction conditions relevant to acetylene conversion. Then, a model for time-on-stream-dependent selectivity focused on situations associated to noticeable acetylene conversion is obtained by the sure-independence-screening-and-sparsifying-operator (SISSO) approach. Our AI approach identifies relationships between the measured catalyst reactivity and only few, key parameters, from 21 measured and calculated bulk, surface, and mesoscopic materials' properties and reaction parameters offered as candidate descriptive parameters. These identified parameters highlight the crucial influence of surface and subsurface carbon and hydrogen on the selectivity towards ethylene formation. Guided by the AI models, new, highly selective bimetallic and trimetallic systems are designed and tested experimentally.

I. INTRODUCTION

Heterogeneous catalysis is governed by a concerted and intricate interplay of several processes that take place at multiple time and length scales.[1–4] The chemical environment that exists during catalyst operation as well as the kinetics of structural changes create configurations of typically unknown local geometries and chemistry. While surface reactions that break and make chemical bonds occur at time scales of picoseconds, they produce species such as carbon, hydrogen, or oxygen, that interact with the material initially placed in the reactor and may induce its local restructuring along with changes in the reactivity in the scale of hours, days, or even longer. Indeed, the catalytic performance of a material, i.e., its ability to more or less selectively convert reactants, typically changes in the beginning of the reaction during the so called *induction period*. During this induction period, the activity and/or the selectivity can drastically change until the surface chemistry reaches a steady state under the applied reaction conditions. At the steady state, a constant catalytic performance is observed with respect to time. At longer reaction times, the catalytic performance might change again, e.g., the activity might decrease. This latter process is known as *deactivation*. The period of time during which the catalytic process is operated is referred to as *time on stream* (t_{OS}).

Modelling the evolution of the reactivity with t_{OS} is crucial for the design of catalytic materials and processes, as it describes the time span during which the material can effectively perform the desired chemical transformation. However, modelling the full catalytic progression at realistic temperatures, pressures and at the long time scales required to capture the t_{OS} evolution of the reactivity from first principles is an inappropriate concept. Thus, systematic approaches for modelling the t_{OS}-dependent catalytic behavior and for the design of improved materials are not well established.

To address the intricacy of heterogeneous catalysis, we put forward a data-centric modelling approach based on consistent experimental data and artificial intelligence (AI).[5, 6] This strategy may capture the multidimensional catalytic progression better than previous computational methods because it targets correlations and does not assume a single underlying physical model. The goal of our approach is to identify the key experimental or calculated physical descriptive parameters that correlate with the reactivity rather than explicitly describe all the underlying physical processes that trigger, just facilitate, or hinder the activity or selectivity. Crucially, the key descriptive parameters are identified from many offered *candidate descriptive parameters*, also called *primary features*. In analogy to genes in biology, the selected key parameters are called "materials genes" of catalysis,[6] since they correlate with the catalytic function of interest. The concept has been previously applied to model the steady-state reactivity in alkane oxidation[6–8], CO oxidation [9], and CO_2 hydrogenation.[10] We note that the relevance and weights of the different processes may be very different for different materials. This questions

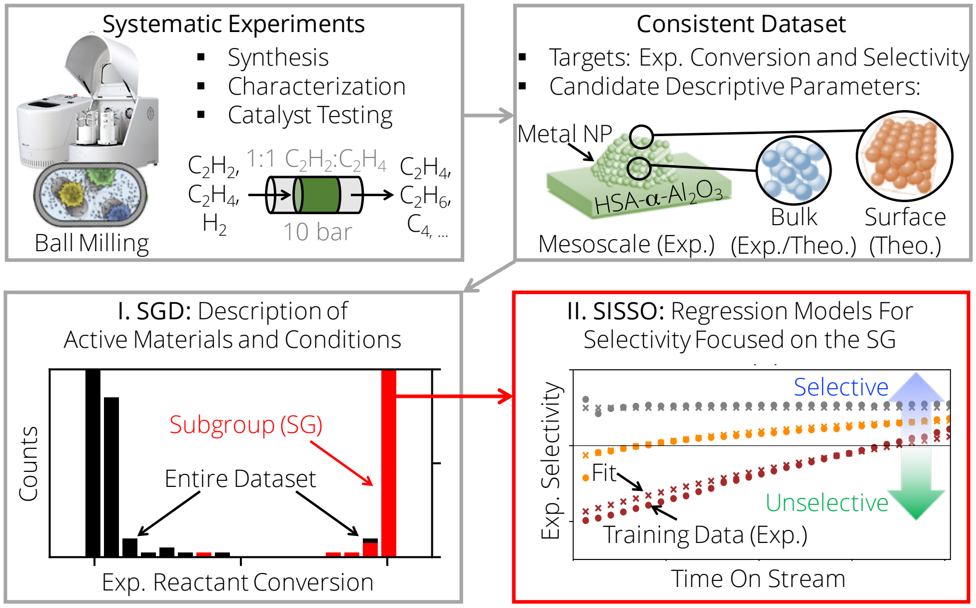

FIG. 1. Focused AI approach for modelling the time-dependent reactivity of metal alloy catalysts for the selective hydrogenation of concentrated acetylene streams. By using systematic experiments (Exp.) and theoretical calculations (Theo.), a consistent dataset containing the measured reactivity and many candidate descriptive parameters (primary features) is created. To this dataset, the subgroup-discovery (SGD) and the sure-independence-screening-and-sparsifying-operator (SISSO) approaches are applied. SGD and SISSO identify the key parameters correlated with activity and selectivity.

the suitability of a single, global AI model to describe all situations.[11]

Here, we introduce a *focused* AI approach to model the evolution of the catalyst reactivity with t_{OS} (Fig. 1). By *focused*, we mean that the AI models describe in greater detail the situations resulting in noticeable (nonzero) activity. Starting from data obtained by systematic experiments and theoretical simulations, the first step of the focused AI approach uses subgroup discovery (SGD)[12–17] to identify descriptions of subsets of materials and applied conditions resulting in noticeable reactant conversion. In the second step, the symbolic-regression sure-independence-screening-and-sparsifying-operator (SISSO) approach[18–22] is used to model the t_{OS}-dependent selectivity focused on the subsets of data identified in the first step. This strategy allows the symbolic-regression model to focus on the mechanisms that govern active catalysts of interest.

Let us illustrate the two-step focused AI approach by modelling the selective hydrogenation of acetylene to ethylene on oxide-supported metal nanoparticle (NP) alloys. In particular, we consider the selective hydrogenation of acetylene in highly concentrated equimolar acteylene-ethylene streams (ca. 14 vol.-% each in hydrogen excess),[23] which would result from a hypothetical electric plasma-assisted methane-to-ethylene conversion plant. The mentioned plasma-assisted process can enable the production of ethylene from natural gas, biogas or hydrogenated CO_2 using the short-term surpluses in electricity from renewable sources.[24] However, acetylene is formed as a by-product in equimolar concentrations to ethylene and it needs to be selectively converted to ethylene in a dedicated, separate downstream process.[25] New catalyst materials are required for this purpose, as the reaction conditions differ significantly from the processes for selective hydrogenation of acetylene traces ($<$2 vol.-%) formed in ethylene streams coming from the steam-cracking of naphtha.[23] As highly reactive compound, acetylene is prone to poison catalysts in the downstream processing of ethylene, e.g., polymerization to polyethylene. Therefore, acetylene has to be removed almost completely from the ethylene streams (down to $<$ 5 ppm).[26–28] We note that studying catalyst reactivity at industrially relevant elevated pressures used for the hydrogenation of concentrated acetylene streams requires specific safety measures due to the high exothermicity of the reaction and the explosive characteristics of the compounds involved. Difficulties to implement those safety measures into analytical techniques impede the use of *in situ* spectroscopy for catalyst characterization under reaction conditions. This prevents the detailed study of the catalyst structure and properties during t_{OS} under realistic conditions.

Recently, we took advantage of a mechanochemical procedure[29–31] to synthesize palladium-silver alloys supported on high-surface-area α-Al_2O_3 (HSA-α-Al_2O_3) as catalysts for the selective hydrogenation of such acetylene streams.[32] The mechanochemical (dry) synthesis of catalytic materials[29–31] presents a higher atom efficiency and lower environmental impact compared to wet methods.[33] Furthermore, the high-energy conditions provided by the mechanochemical treatment can provide materials with improved properties compared to the materials obtained by traditional methods, e.g., materials with higher surface area.[34] Thus, mechanochemistry is promising for obtaining new, high-performant catalysts. The synthesized palladium-silver materials display high selectivity towards ethylene with significantly higher stability with t_{OS} under the harsh reaction conditions compared to materials synthesized by wet impregnation.[32] However, systematic approaches to navigate the immense space of possible materials that can be made mechanochemically are not yet available. Here, we combine the advantages of the mechanochemical synthesis with focused AI in order to efficiently design improved palladium-based bimetallic and trimetallic systems for the selective hydrogenation of concentrated acetylene streams.

II. METHODS

A. Materials Synthesis and Catalytic Testing

The materials were synthesized mechanochemically via ball milling according to previously published procedures.[29–31, 35] The synthesis details are available in the Electronic Supplementary Information (ESI). Metal-containing compounds in powder were milled in the presence of γ-AlOOH (bohemite), leading to the formation of metal NPs supported on HSA-α-Al_2O_3 with a relative amount of metal with respect to support, in weight, of ca. 1%, noted 1 wt%. The synthesized materials were subjected to a thermal activation procedure consisting of a reduction step with H_2 and an annealing step.

The catalytic tests were performed in a continuous-flow fixed-bed reactor at 10 bar. The oven containing the reactor was heated to a set temperature (T_{oven}) of 50, 100, or 150 °C. The applied feed contains a H_2:C_2H_2:C_2H_4 ratio of 5:1:1. The acetylene conversion ($X_{\mathrm{C_2H_2}}$) was evaluated as

$$X_{\mathrm{C_2H_2}} = 1 - \frac{n_{\mathrm{C_2H_2,out}}}{n_{\mathrm{C_2H_2,in}}}. \tag{1}$$

Here, $n_{\mathrm{C_2H_2,out}}$ and $n_{\mathrm{C_2H_2,in}}$ are molar flow rates of acetylene leaving and entering the reactor, respectively. The values of $X_{\mathrm{C_2H_2}}$ are in the range [0,1]. The values of zero and one correspond, respectively, to inactive materials and reaction conditions, and to 100% (or full) acetylene conversion. The selectivity towards the formation of ethylene ($S_{\mathrm{C_2H_4}}$) was evaluated as

$$S_{\mathrm{C_2H_4}} = \frac{n_{\mathrm{C_2H_4,out}} - n_{\mathrm{C_2H_4,in}}}{n_{\mathrm{C_2H_2,in}} - n_{\mathrm{C_2H_2,out}}}. \tag{2}$$

In Eq. 2, $n_{\mathrm{C_2H_4,out}}$ and $n_{\mathrm{C_2H_4,in}}$ are molar flow rates of ethylene leaving and entering the reactor, respectively. Note that if $X_{\mathrm{C_2H_2}} = 0$, the denominator of Eq. 2 is equal to zero and $S_{\mathrm{C_2H_4}} = 0$ by convention. The values of $S_{\mathrm{C_2H_4}}$ are in the range [-1,1]. Ethylene is both a product of the acetylene selective hydrogenation reaction and part of the reaction feed. $S_{\mathrm{C_2H_4}} < 0$ indicates that more ethylene (from the feed) was consumed than formed from acetylene, with $S_{\mathrm{C_2H_4}} = -1$ corresponding to the full consumption of the ethylene of the feed. $S_{\mathrm{C_2H_4}} > 0$ indicate selective materials and conditions. The ideal, desired behavior corresponds to $S_{\mathrm{C_2H_4}} = 1$. In addition to the formation of the desired ethylene, the undesirable formation of the total-hydrogenation product ethane and the formation of oligomeric species (e.g., C_4) were also quantified (see ESI for details). The reactivity was measured at each 13.5 minutes (min) during t_{OS}. Additional details on the catalytic tests are provided in ESI and elsewhere.[32, 36–39]

B. Experimental and Calculated Candidate Descriptive Parameters

In order to model the reactivity via AI, we collected candidate descriptive physical parameters characterizing the materials, reaction conditions, and potentially relevant processes governing acetylene hydrogenation. Four parameters were obtained from the experimental characterization of the materials by energy dispersive X-ray analysis in a scanning electron microscope (EDX-SEM), high-resolution transmission-electron microscopy (TEM), and N_2 physisorption, such as the mean particle diameter (D_μ) and the specific surface area (s_{BET}).

Additionally, we considered elemental properties of the metals in order to construct candidate descriptive parameters reflecting the chemistry of the bulk of the metal NPs. These elemental properties are atomic (free-atom) properties such as ionization potential (IP) and bulk properties of solids such as closest interatomic distance (d_{closest}) and cohesive energy (E_{coh}). The elemental properties were converted into materials-specific parameters by taking the composition average

$$\overline{\phi} = \sum x_i \phi_i. \tag{3}$$

In Eq. 3, ϕ is an elemental property and x_i are the nominal molar fractions of each metal in the materials' composition, e.g., $x_{\mathrm{Pd}} = 1/10$ and $x_{\mathrm{Ag}} = 9/10$ for the Pd_1Ag_9 material.

We also constructed candidate descriptive parameters reflecting the properties of the surfaces of the metal NPs and their interaction with the reaction environment. For this purpose, we considered parameters calculated by theory using density functional theory under the generalized gradient approximation (DFT-GGA) on low-index model surfaces. These model surfaces are by no means representative of the complex structural motifs that might be formed on the catalysts under reaction conditions. However, our AI analysis will combine these basic parameters with other bulk and mesoscale materials' parameters according to linear and nonlinear operators to appropriately describe the reactivity. We included the energy of the d-band center (ϵ_d) as a property of the clean surface, as it correlates with the adsorption energies.[40] Besides, we considered eight parameters that describe the interaction of carbon and hydrogen with the surface and subsurface, as these species were suggested as crucial in the selective hydrogenation of diluted acetylene streams on palladium-based catalysts.[41–43] Examples of such parameters are the binding energy of subsurface hydrogen and carbon ($E^{\rm sub}_{\rm b,H}$ and $E^{\rm sub}_{\rm b,C}$, respectively).[44, 45] The surface- and metal-dependent properties were also converted into materials-specific parameters via Eq. 3. Finally, we included $t_{\rm OS}$ and $T_{\rm oven}$ as parameters related to the applied reaction conditions. In total, 21 candidate descriptive parameters were collected (see full list in Table S1 as well details on the experimental or theoretical methods used to obtain them).

C. Focused AI Approach

The two-step focused AI modelling approach takes into account two different materials design criteria or targets, (i) the acetylene conversion and (ii) the selectivity, e.g., towards the formation of ethylene. In the first step, subgroup discovery (SGD)[12, 13, 16] is applied to identify subsets of materials and reaction conditions that exhibit noticeable acetylene conversion. In the second step, the symbolic-regression SISSO approach[18] is used to model the $t_{\rm OS}$-dependent selectivity towards ethylene, ethane and C_4 products, for the subset of materials and conditions identified in the first step. Therefore, the SGD model will indicate whether a given material and condition are associated with noticeable conversion. The SISSO model, in turn, will quantitatively describe the evolution of the selectivity with $t_{\rm OS}$ for the active materials and conditions. Thus, the SISSO models for selectivity do not attempt to describe all materials and reaction conditions simultaneously, but they focus on the situations that present noticeable acetylene conversion. The SGD approach could be directly applied to identify descriptions of materials and reaction conditions resulting in the desired selectivity values, e.g., $S_{\rm C_2H_4} > 0$, in the first step of the AI approach. However, such strategy would not provide a detailed description of the change in of selectivity with $t_{\rm OS}$, which for some materials is dramatic, e.g., $S_{\rm C_2H_4} < 0$ at reaction start and $S_{\rm C_2H_4} > 0$ at long $t_{\rm OS}$.

The SGD approach[12, 13] identifies partitions of the data space associated with outstanding distributions of a given target of interest, here the acetylene conversion. SGD starts by creating statements about the candidate descriptive parameters that are only satisfied by a portion of the dataset. Such statements are, for instance, inequalities constraining the values of the descriptive parameters to some minimum or maximum thresholds to be determined during the analysis. Then, SGD identifies conjunctions of statements that result in subselections of data (SGs) that maximize an objective function. This objective function specifies how outstanding the SGs are with respect to the entire dataset based on their sizes and on the distributions of the target values in the SG and in the entire dataset. Here, we use the objective function

$$Q(SG,\widetilde{P}) = \frac{s(SG)}{s(\widetilde{P})} * \frac{X_\sigma(SG) - X_\sigma(\widetilde{P})}{X_\sigma(\widetilde{P})}, \qquad (4)$$

where $s(SG)$ and $s(\widetilde{P})$ are the sizes of the SG and of the entire dataset $\widetilde{P}$, respectively, and $X_\sigma(SG)$ and $X_\sigma(P)$ are the standard deviation of the distribution of the target in the SG and in the entire dataset, respectively. The so-called *standard-deviation-reduction* utility function, $\frac{X_\sigma(SG)-X_\sigma(\widetilde{P})}{X_\sigma(\widetilde{P})}$, favors the identification of SGs associated with narrow distributions of target values. The outcome of the SGD analysis is the description of the SG that maximizes the objective function. This description constrains the values of only few, key parameters, out of the many initially offered ones and might be referred to as *SG rules*. Importantly, the SG description focuses on the behavior of the data points that belong to the SG. In this sense, the SG rules are local, not global.

The SISSO approach[18] identifies (nonlinear) analytical expressions that describe a given data set. Starting from the candidate descriptive parameters, SISSO builds an immense pool of analytical expressions (e.g., containing billions of expressions) by iteratively applying mathematical operators such as addition, subtraction, exponential, and more to the primary features and to previously generated expressions. Then, by compressed sensing[46, 47] SISSO identifies the few expressions that, combined by weighting coefficients, best correlate with a given target of interest, here the selectivity. Typically, only few key descriptive parameters, out of all initially offered candidates, appear in the models identified by SISSO. The predictive performance and optimal complexity of the SISSO models were assessed by a nested five-fold cross-validation scheme. By determining SISSO model hyperparameters and thus the model complexity via cross validation, we prevent the identification of models that fit well the training set but which are not generalizable (overfitting). Additional details on the SGD and SISSO approaches are described in ESI and in references[6, 18, 22].

III. RESULTS AND DISCUSSION

A. Synthesized Materials and Their Reactivity in Acetylene Selective Hydrogenation

Twelve materials constituted of HSA-α-Al_2O_3-supported metal NPs were synthesized via the mechanochemical approach.[29–32] These are three monometalic systems (Ag, Au, and Cu) and nine bimetallic alloys with nominal molar ratios 1:1, 1:5, and 1:9, namely Pd_1Ag_1, Pd_1Ag_5, Pd_1Ag_9, Pd_1Au_1,

Pd_1Au_5, Pd_1Au_9, Pd_1Cu_1, Pd_1Cu_5, Pd_1Cu_9. These materials were tested in the selective hydrogenation of concentrated acetylene streams in the presence of equimolar amounts of ethylene.

The distributions of measured acetylene conversions ($X_{\mathrm{C_2H_2}}$) and ethylene selectivities ($S_{\mathrm{C_2H_4}}$) are shown in black in Fig. 2A and B, respectively. These distributions correspond to measurements of the 12 materials under three T_{oven} and associated to a number of t_{OS} values between 0 and 405 min. In total, 1,076 measurements were performed. The distribution of measured $X_{\mathrm{C_2H_2}}$ values is highly concentrated around the values of 0.0 and 1.0, indicating that most of the materials and reaction conditions either result in inactive systems or in full conversion of acetylene, respectively. The high exothermicity of the reaction at such harsh reaction conditions impedes active catalysts to stabilize at acetylene conversion levels below 100 % without sophisticated cooling systems. We refer at this point to the work from Van Heerden on stationary states in autothermic processes.[48, 49] For this reason, we decided to analyze catalyst performance for all catalysts at stable full conversion levels even though this situation is not ideal in consecutive hydrogenation reactions such as the acetylene hydrogenation reaction. The distribution of measured $S_{\mathrm{C_2H_4}}$ values is in the approximate range [-1.0, 0.75], reflecting that the tested materials and conditions result in diverse behaviors ranging from totally unselective to highly selective hydrogenation of acetylene. A significant number of measurements results in $S_{\mathrm{C_2H_4}}$ close to -1.0, which correspond to total hydrogenation of both acetylene and co-fed ethylene to ethane. Additionally, many systems are associated to $S_{\mathrm{C_2H_4}} = 0$, which indicates either equivalent rates of co-fed ethylene hydrogenation and ethylene formation from acetylene or no activity ($X_{\mathrm{C_2H_2}} = 0$).

B. Identification of Subgroups of Materials and Reaction Conditions Resulting in Noticeable Acetylene Conversion

In order to identify descriptions of materials and reaction conditions that provide noticeable acetylene conversion, we applied SGD. We searched for SGs of catalysts and conditions that present a narrow distribution of $X_{\mathrm{C_2H_2}}$. In the present dataset, the majority of situations associated with noticeable activity correspond to $X_{\mathrm{C_2H_2}}$ values close to 1.0 (Fig. 2A). Removing data points with $X_{\mathrm{C_2H_2}} = 0$ and $S = 0$ from the SISSO modelling prevents that the uninteresting yet numerous situations associated to zero activity dominate the selection of models for selectivity, as the root-mean-squared loss (objective) function of SISSO reflects the average performance across the entire training data. Thus, the loss function would be strongly affected by the data points with $X_{\mathrm{C_2H_2}} = 0$ and $S = 0$. Besides, by considering a narrow, nonzero conversion range, a proper comparison of selectivity across materials and reaction conditions is possible (see next section).

The identified SG contains 539 data points (ca. 50% of the dataset) and the distribution of $X_{\mathrm{C_2H_2}}$ for the data points that belong to this SG is shown in red in Fig. 2A. Most of the materials and reaction conditions associated to high, close-to-one conversion values are found within the identified SG. The distribution of $S_{\mathrm{C_2H_4}}$ for the data points that belong to this SG is shown in red in Fig. 2B. The situations associated with $X_{\mathrm{C_2H_2}} = 0$ and $S_{\mathrm{C_2H_4}} = 0$ that were present in the dataset (black bar at $S_{\mathrm{C_2H_4}} = 0$ in Fig. 2B) are not part of the SG. The identified SG is described by the rules σ_X^{SG}. They constrain the values of three descriptive parameters:

$$\sigma_X^{\mathrm{SG}} \equiv T_{\mathrm{oven}} \geq 75^{\circ}\mathrm{C} \wedge \overline{d_{\mathrm{closest}}} \geq 2.57\text{Å} \wedge \overline{E_{\mathrm{b,C}}^{\mathrm{sub}}} \geq 4.67\mathrm{eV}. \tag{5}$$

In Eq. 5, T_{oven} is the temperature of the oven in which the reactor is placed, $\overline{d_{\mathrm{closest}}}$ is the composition-averaged closest interatomic bulk distance, and $\overline{E_{\mathrm{b,C}}^{\mathrm{sub}}}$ is the composition-averaged binding energy of subsurface carbon. The symbol ∧ corresponds to the "AND" operator. According to the SG rules, the full acetylene conversion can be associated with the temperatures of 100 and 150 °C and with systems presenting $\overline{d_{\mathrm{closest}}}$ and $\overline{E_{\mathrm{b,C}}^{\mathrm{sub}}}$ above certain thresholds. The relevance of the parameters $\overline{d_{\mathrm{closest}}}$ and $\overline{E_{\mathrm{b,C}}^{\mathrm{sub}}}$ indicates that bulk as well as surface properties of the metal NPs determine high conversion levels. In particular, the parameter $\overline{E_{\mathrm{b,C}}^{\mathrm{sub}}}$ highlights that the activity depends on the subsurface interaction with carbon. The candidate descriptive parameters obtained from the experimental characterization of the materials and the parameters reflecting the properties of the clean surface and the interaction of the surfaces and subsurfaces of the metal NPs with hydrogen do not appear in the description of the identified SG. Finally, we also note that the candidate descriptive parameter t_{OS}, also offered in the SGD analysis, is not selected in Eq. 5. Indeed, we do not observe significant variation of $X_{\mathrm{C_2H_2}}$ as a function of t_{OS} (see ESI for details). The 12 materials of the dataset are represented in the coordinates of the key materials-dependent parameters identified by SGD ($\overline{d_{\mathrm{closest}}}$ and $\overline{E_{\mathrm{b,C}}^{\mathrm{sub}}}$) in Fig. 2C. In this plot, the bimetallic alloys, which are part of the SG, are shown as red circles. The monometallic systems, which are not part of the SG, are shown as black circles. The SG rules are represented by the dashed red lines and arrows, and the SG region is highlighted in red.

While in this application the SG corresponds to high catalytic activity, it should be noted that the SGD approach could be used to identify descriptions of data points associated with moderate or low activity in other systems. This is important because highly active materials and conditions are not necessarily those displaying the desired selectivity, e.g. in consecutive reactions.[6, 8]

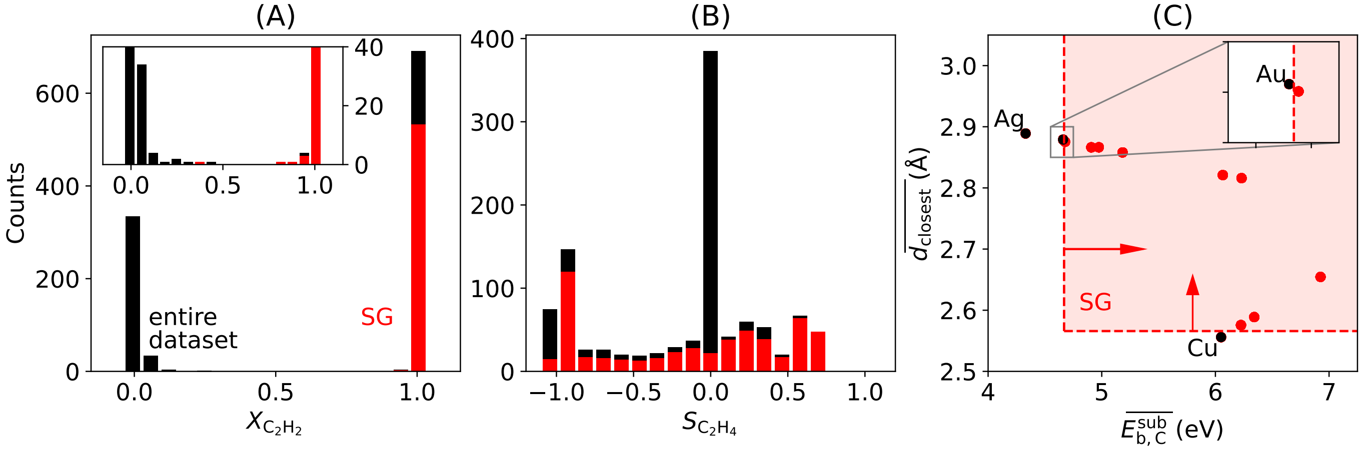

FIG. 2. Identification of subgroup rules describing materials and reaction conditions resulting in acetylene conversion in the catalytic hydrogenation of concentrated acetylene using high-surface-area Al_2O_3-supported metal alloys synthesized mechanochemically. (A): Distribution of acetylene conversion ($X_{C_2H_2}$, Eq. 1) in the entire dataset and in the identified SG. The inset shows in detail the regions of the plot associated to low y-axis values. The height of the inset box is 40 counts. (B): Distribution of ethylene selectivity ($S_{C_2H_4}$, Eq. 2) in the entire dataset and in the identified SG. (C): The SG rules constrain the value of the key descriptive parameters composition-averaged closest interatomic distance ($\overline{d_{\mathrm{closest}}}$) and subsurface carbon binding energy ($\overline{E^{\mathrm{sub}}_{\mathrm{b,C}}}$), see full set of constraints in Eq. 5.

C. Identification of Symbolic-Regression Models for Time-Dependent Selectivity Focused on Noticeable Acetylene Conversion

We now analyze the selectivity of the materials and reaction conditions that present high activity and belong to the identified SG. The evolution of measured $S_{C_2H_4}$ values with t_{OS} for the materials in the SG are shown in Fig. 3A and B for T_{oven} 100 and 150 °C, respectively. Some materials such as Pd_1Ag_9 and Pd_1Ag_5 present a near-constant $S_{C_2H_4}$ for $t_{\mathrm{OS}} > 30$ min. This indicates that for these materials, the steady state is reached in less than 30 min. However, other materials develop an improved selectivity as t_{OS} increases. In particular, the Pd_1Cu_9 material presents $S_{C_2H_4} = -1$ (total hydrogenation catalyst) at $t_{\mathrm{OS}} < 30$ min, and the $S_{C_2H_4}$ value becomes positive (selective catalyst) after ca. 180 min in the case of $T_{\mathrm{oven}} = 150°$ - see Fig. 3B, square brown markers. Thus, for some catalysts, the compound initially placed in the reactor is undergoing significant modifications. The induction period for reaching optimal performance and the steady state is in the range of hours. We stress that all materials shown in Fig. 3 are active towards acetylene conversion (with $X_{C_2H_2} \approx 1$) since the start of the reaction at $t_{\mathrm{OS}} = 0$.

By using only the data points belonging to the SG associated to high acetylene conversion (539 data points), we trained SISSO models for $S_{C_2H_4}$. The mean prediction (test) error evaluated using five-fold nested cross validation is 0.105 for SISSO models with the optimal complexity. Such value is significantly lower than the standard deviation of $S_{C_2H_4}$ in the training data set (0.600). This shows that the identified models describe unseen data with good accuracy. The fit of the best SISSO models to the data, shown in Fig. 3C and D for T_{oven} 100 and 150 °C, respectively, indicates that the identified expression (Eq. 6) is able to capture the selectivity trends both across materials and with t_{OS}. In particular, the selectivity shift of the Pd_1Cu_9 material (Fig. 3C and D, square brown markers) with t_{OS} is well captured by the AI model.

$$\begin{aligned} S^{\mathrm{SISSO}}_{C_2H_4} =& c_0^{C_2H_4}(T_{\mathrm{oven}}) \\ &+ c_1^{C_2H_4}(T_{\mathrm{oven}}) * \frac{t_{\mathrm{OS}}}{\overline{\Delta W^{\mathrm{surf}}_{\mathrm{H}}} * |\overline{\epsilon_d} - \overline{E^{\mathrm{surf}}_{\mathrm{b,H}}}|} \\ &+ c_2^{C_2H_4}(T_{\mathrm{oven}}) * \overline{E^{\mathrm{sub}}_{\mathrm{d,C}}} * w_{\mathrm{metal}} * D_{\mu} \\ &+ c_3^{C_2H_4}(T_{\mathrm{oven}}) * e^{\{8.37*10^{-4}\,\frac{\mathrm{eV}}{\mathrm{min}}(\frac{t_{\mathrm{OS}}+359\ \mathrm{min}}{\overline{E^{\mathrm{sub}}_{\mathrm{b,H}}}+1.38\ \mathrm{eV}})\}} \end{aligned} \tag{6}$$

In Eq. 6, the coefficients are denoted $c_i^{C_2H_4}(T_{\mathrm{oven}})$ to highlight that they correspond to different fitted values for the two different T_{oven} (see ESI for details). Firstly, we note that the model expression contains the key parameter t_{OS}, required to describe the ethylene selectivity time dependency. Additionally, the following materials-dependent theoretical key parameters appear in Eq. 6: work-function shift with hydrogen adsorption ($\overline{\Delta W^{\mathrm{surf}}_{\mathrm{H}}}$), d-band center ($\overline{\epsilon_d}$), binding energy of hydrogen ($\overline{E^{\mathrm{surf}}_{\mathrm{b,H}}}$), deformation energy of subsurface carbon ($\overline{E^{\mathrm{sub}}_{\mathrm{d,C}}}$), and binding energy of subsurface hydrogen ($\overline{E^{\mathrm{sub}}_{\mathrm{b,H}}}$). Finally, the total metal loading (w_{metal}) and mean value of particle size (D_{μ}), measured prior to the catalytic test, are also identified as key experimental parameters. The relevance of the microscopic parameters $\overline{\epsilon_d}$, $\overline{\Delta W^{\mathrm{surf}}_{\mathrm{H}}}$, $\overline{E^{\mathrm{surf}}_{\mathrm{b,H}}}$, $\overline{E^{\mathrm{sub}}_{\mathrm{d,C}}}$,

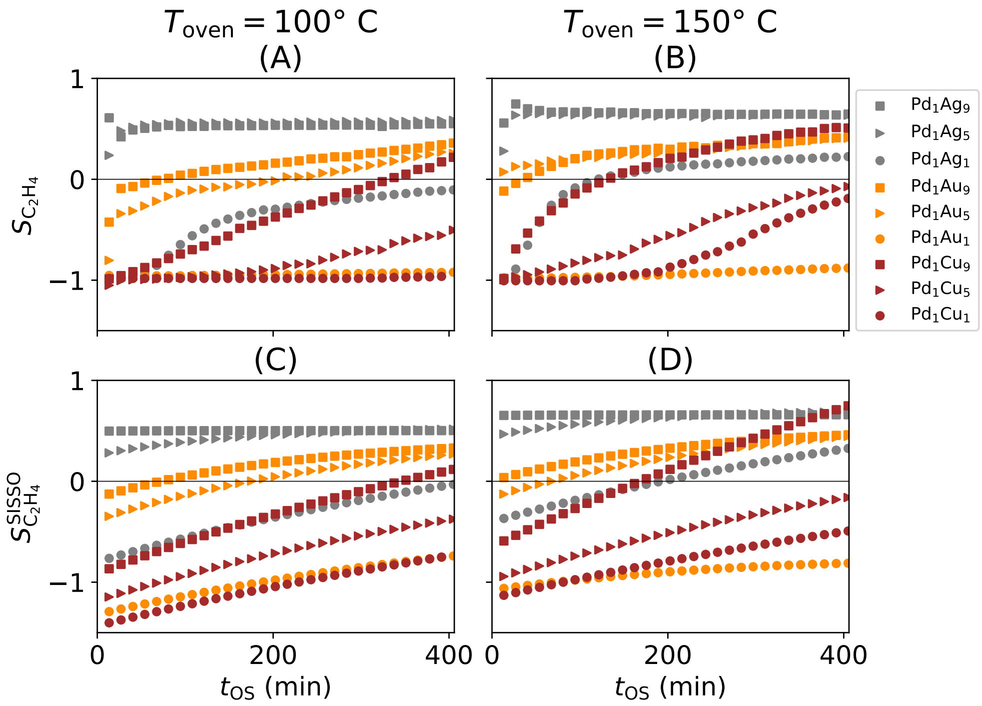

FIG. 3. Measured (top row) and SISSO-modelled (bottom row) selectivity towards ethylene formation ($S_{\mathrm{C_2H_4}}$, Eq. 2). The two columns display results at different oven temperatures T_{oven} 100 and 150 °C, respectively. $S_{\mathrm{C_2H_4}} < 0$ indicates that more ethylene (from the feed) was consumed than formed from acetylene. $S_{\mathrm{C_2H_4}} > 0$ indicate that more ethylene was formed from acetylene than consumed (selective materials and conditions).

and $\overline{E^{\mathrm{sub}}_{\mathrm{b,H}}}$ indicates that the surface properties, including the interaction of surface and subsurface with carbon and hydrogen, are crucial for the selectivity towards ethylene. This might be related to the formation of surface carbides and the availability of surface hydrogen.[41] Indeed, the formation of $\mathrm{Pd}_x\mathrm{C}_y$ surface layers has been shown to lower the desorption barrier of ethylene while the low availability of surface hydrogen atoms can prevent the undesired further hydrogenation of ethylene to ethane, both enhancing $S_{\mathrm{C_2H_4}}$. [50] Additionally, the relevance of D_μ and w_{metal} could be related to the type and relative amount of surface sites available, which depend on the distribution of NP sizes and on the metal loading. Interestingly, the t_{OS}-independent term of Eq. 2 contains the parameters $\overline{E^{\mathrm{sub}}_{\mathrm{d,C}}}$, w_{metal}, and D_μ, whereas the t_{OS}-dependant terms of Eq. 2 contain the parameters $\overline{\epsilon_d}$, $\overline{\Delta W^{\mathrm{surf}}_{\mathrm{H}}}$, $\overline{E^{\mathrm{surf}}_{\mathrm{b,H}}}$, and $\overline{E^{\mathrm{sub}}_{\mathrm{b,H}}}$. Therefore, the stability of the subsurface carbon and the loading and size of metal NPs mainly determine the initial value of ethylene selectivity at $t_{\mathrm{OS}} = 0$ min, while the stability of surface and subsurface hydrogen is the key factor modulating the evolution of the selectivity with t_{OS}. No elemental or metal bulk candidate descriptive parameters are selected by SISSO in Eq. 6. This suggests that the ethylene selectivity is governed by processes related to the NP surface and subsurface rather than its bulk. We note that many analytical expressions considered by SISSO may provide a similar description compared to Eq. 6. Thus, the expressions and operators themselves might not have physical meaning. However, the parameters entering the expressions do have a physical meaning, as they correlate with relevant underlying physical processes. We have also trained SISSO models for the selectivity towards ethane. These results are discussed in the ESI.

In addition to ethylene and ethane, $\mathrm{C_4}$ products such as butane, 1-butene, 2-cis-butene, and 2-trans-butene, are formed during the acetylene hydrogenation. The selectivity towards $\mathrm{C_4}$ products is lower than the selectivity towards ethylene or ethane ($S_{\mathrm{C_4}} < 0.2$). However, the formation of $\mathrm{C_4}$ products indicates the tendency of carbon-carbon bond formation, and thus the tendency of forming acetylene oligomers, often referred to as *green oil*. These oligomers can accumulate on the surface of the material leading to the loss of activity (deactivation) typically at long t_{OS}.[32] In order to get insights on the processes that lead to oligomer formation and thus to catalyst deactivation, we also trained a SISSO model for $S_{\mathrm{C_4}}$ by using the 539 data points belonging to the SG associated to high acetylene conversion. This model is noted in Eq. 7 (see further details in ESI).

$$\begin{aligned} S^{\mathrm{SISSO}}_{\mathrm{C}_4} =& c_0^{\mathrm{C}_4}(T_{\mathrm{oven}}) \\ &+ c_1^{\mathrm{C}_4}(T_{\mathrm{oven}}) * \frac{\ln\{1.0\frac{1}{\mathrm{min}} * (t_{\mathrm{OS}} - 4.05\ \mathrm{min})\}}{\overline{E^{\mathrm{surf}}_{\mathrm{b,H}}} * s_{\mathrm{BET}}} \\ &+ c_2^{\mathrm{C}_4}(T_{\mathrm{oven}}) * \frac{(\overline{\delta^{\mathrm{sub}}_{\mathrm{C}}})^6}{\overline{\epsilon_d} - \overline{E^{\mathrm{surf}}_{\mathrm{b,H}}}} \\ &+ c_3^{\mathrm{C}_4}(T_{\mathrm{oven}}) * \frac{\overline{d_{\mathrm{closest}}} * (\overline{E^{\mathrm{sub}}_{\mathrm{d,C}}} + \overline{EA})}{\overline{EA}} \end{aligned} \tag{7}$$

The key parameters identified in Eq. 7 are t_{OS}, $\overline{\epsilon_d}$, $\overline{E^{\mathrm{surf}}_{\mathrm{b,H}}}$, $\overline{E^{\mathrm{sub}}_{\mathrm{d,C}}}$, composition-averaged distance expansion due to subsurface carbon ($\overline{\delta^{\mathrm{sub}}_{\mathrm{C}}}$), $\overline{d_{\mathrm{closest}}}$, composition-averaged electron affinity ($\overline{EA}$), and specific surface area (s_{BET}). The microscopic parameters $\overline{\epsilon_d}$, $\overline{E^{\mathrm{surf}}_{\mathrm{b,H}}}$, $\overline{E^{\mathrm{sub}}_{\mathrm{d,C}}}$, and $\overline{\delta^{\mathrm{sub}}_{\mathrm{C}}}$, indicate that (sub)surface processes, including the participation of carbon and hydrogen, are crucial for the formation of C_4 molecules. Additionally, the parameters $\overline{d_{\mathrm{closest}}}$ and $\overline{EA}$ show that properties of the bulk of NP are important. Finally, the key parameter s_{BET} reflects that the surface area also impacts the formation of C_4 products. The relevance of the (sub)surface parameters $\overline{\epsilon_d}$, $\overline{E^{\mathrm{surf}}_{\mathrm{b,H}}}$, $\overline{E^{\mathrm{sub}}_{\mathrm{d,C}}}$, and $\overline{\delta^{\mathrm{sub}}_{\mathrm{C}}}$ can be related to the influence of surface-carbon binding and hydrogen availability on the reaction pathways that lead to C_4 products. Indeed, it had been argued that surface sites that provide strong palladium-carbon binding are more prone to oligomerize acetylene and other surface intermediates during the reaction.[50] The relevance of the bulk parameters $\overline{d_{\mathrm{closest}}}$, $\overline{EA}$, in turn, could be related to the de-alloying of the NPs.[51, 52] This de-alloying process in palladium-based materials, for instance, results in the creation of palladium-rich surface sites, which could more efficiently oligomerize acetylene compared to sites containing both palladium and other metals such as silver.[50] Indeed, the X-ray diffraction (XRD) analysis of some of the palladium-silver alloys after the reaction[32] indicates the segregation of palladium to the surface. In case of a fixed particle size, s_{BET} is correlated with the distances between NPs. The relevance of this parameter could point at the importance of average distances between particles in the oligomerisation side reaction. However, the average particle sizes in the investigated data set vary between 3 to 9 nm. Thus, this interpretation has to be handled with care. Interestingly, the time-dependent term of Eq. 7, which captures the S_{C_4} increase with t_{OS} (see Fig. S8), contains the parameters $\overline{E^{\mathrm{surf}}_{\mathrm{b,H}}}$ and s_{BET}. Thus, the availability of hydrogen as well as underlying processes related to the specific surface area modulate the evolution of the formation of C_4 molecules over time and are likely associated to catalyst deactivation.

While some of the microscopic, (sub)surface-related parameters of the $S^{\mathrm{SISSO}}_{\mathrm{C}_4}$ model of Eq. 7 also appear in the $S^{\mathrm{SISSO}}_{\mathrm{C}_2\mathrm{H}_4}$ model of Eq. 6, the bulk parameters and s_{BET} only appear in the expression for the selectivity towards C_4. This indicates that, in addition to surface processes, bulk phenomena related to the properties of the NP volume and the distance between NPs could impact the formation of C_4 molecules, and likely that of the oligomers.

By manually selecting data points associated with $X_{\mathrm{C}_2\mathrm{H}_2} \approx 1$, one could obtain similar data sets compared to the data sets used to train the SISSO models for selectivity discussed in this section. However, such an approach would not indicate whether a new material, i.e., a material that was not tested in catalysis, is active. Conversely, the SG rules can be used to predict which new materials are active for the acetylene hydrogenation, as it will be shown in the next section.

D. Exploitation of the AI Models to Design Bimetallic and Trimetallic Systems and Experimental Verification

Let us now exploit the SGD and SISSO models trained using a data set containing palladium alloys with silver, gold, and copper to design bimetallic systems containing a wider range of stoichiometries as well as other chemical elements. We focus on the hypothetical materials $\mathrm{Pd}_{(1-x_M)}M_{x_M}$, where $M=$ Fe, Co, Ni, Cu, Ru, Rh, Ag, Pt, Au. Note that not all of these hypothetical stoichiometries might result in solid solutions. Besides, not all of these materials might be synthesizable via the mechanochemical approach. Values of x_M in the range [50,90] % are analyzed. The bimetallic materials with x_M 45, 55, 65, 75, 85, 90% are represented as grey circles in the $\overline{d_{\mathrm{closest}}}$ vs. $\overline{E^{\mathrm{sub}}_{\mathrm{b,C}}}$ plot of Fig. 4A. In this plot, the monometallic systems are represented as black circles and the SG rules describing materials associated to high acetylene conversion are displayed as red dashed lines. The red shaded area corresponds to the materials predicted to have high acetylene conversion. The systems containing platinum and ruthenium are predicted to fully convert acetylene irrespective of x_M. This is consistent with the fact that these metals are traditional hydrogenation catalysts.[51, 53] Conversely, according to the SG rules, the systems based on iron, cobalt, nickel, copper, rhodium, silver, and gold only present high acetylene conversion below a certain x_M value, i.e., the amount of palladium has to be higher than a certain threshold. While nickel and rhodium are reported to be active for the hydrogenation of diluted acetylene,[26, 28] they have not yet been tested in the high-concentration, high-pressure conditions in order to verify these predictions.

The ethylene selectivity values for the bimetallic systems $\mathrm{Pd}_{(1-x_M)}M_{x_M}$ were predicted based on Eq. 6. To make these predictions, all parameters appearing in Eq. 6 need to be specified. However, the experimental parameters w_{metal} and D_μ are unknown for materials that were not yet synthesized and characterized. In order to make the predictions, we assume w_{metal} and D_μ to be equal

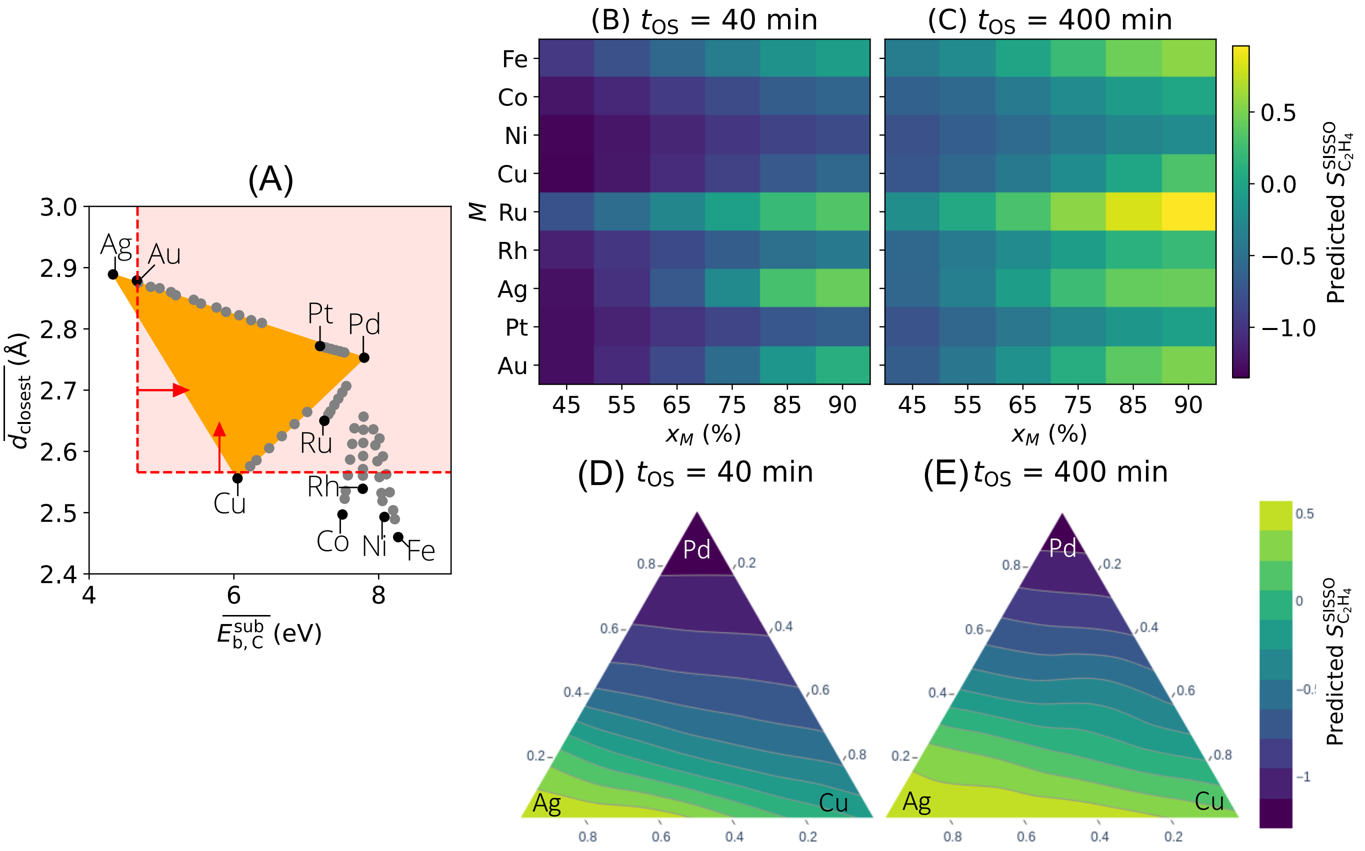

FIG. 4. The SG rules and SISSO models derived based on palladium-, silver-, gold-, and cooper-based materials synthesized by ball milling are used to design new bi- and tri-metallic palladium-based systems for the catalytic hydrogenation of concentrated acetylene. (A): The SG rules on composition-averaged closest interatomic distance ($\overline{d_{closest}}$) and subsurface carbon binding energy ($\overline{E^{sub}_{b,C}}$) identify alloy compositions that fully convert acetylene. (B) and (C): Ethylene selectivity predicted by the SISSO model ($S^{SISSO}_{C_2H_4}$) for bimetallic systems ($Pd_{(1-x_M)}M_{x_M}$, with $M=$ Fe, Co, Ni, Cu, Ru, Rh, Ag, Pt, Au) at two t_{OS}s. (D) and (E): Ethylene selectivity predicted by the SISSO model ($S^{SISSO}_{C_2H_4}$) for trimetallic systems $Pd_xAg_yCu_{1-x-y}$ at two t_{OS}s. The selectivity predictions are made for $T_{oven} = 150°C$.

to the mean values of these parameters in the training set, i.e., 0.012 and 5.67 nm, respectively. Thus, the predictions of Fig. 4B and C might vary if the systems exhibit w_{metal} and D_μ values significantly different from the mean values of the training set.

The color scale in the (x_M,M) composition matrix of Fig. 4B and C reflects the $S^{SISSO}_{C_2H_4}$ values at $T_{oven} = 150°C$ for t_{OS} 40 and 400 min, respectively. For $t_{OS} = 40$ min, the materials containing high amounts of ruthenium, silver, and gold present $S^{SISSO}_{C_2H_4} > 0$. For $t_{OS} = 400$ min, the ethylene selective increases, and some compositions with high x_M based on iron, cobalt, and copper also reach $S^{SISSO}_{C_2H_4} > 0$. However, iron-, cobalt-, copper-, silver-, and gold-based materials with high x_M do not satisfy the SG rules for high $X_{C_2H_2}$ shown in Fig. 4A. Thus, the AI models indicate that bimetallic systems based on these metals could be selective at the expense of a lower (possibly close-to-zero) activity. Among the considered bimetallic systems, the ruthenium-based materials with high x_M are predicted to be the most selective ones. For instance, the bimetallic material with $x_{Ru} = 90\%$ is associated to $S^{SISSO}_{C_2H_4} = 0.95$ for $t_{OS} = 400$ min. This material is also predicted to be highly active according to the SG rules shown in Fig. 4A. However, Ru-Pd systems are not expected to be miscible at any composition according to the phase diagram.[54] Alternatively, the SISSO model indicates that silver-based materials with x_M higher than 90% might be even more selective than the material with the highest selectivity in the training set, Pd_1Ag_9.

In order to test the predictions of the AI models, we synthesized, characterized by TEM and SEM-EDX, and tested in catalysis the new Pd_1Ag_{12} and Pd_1Ag_{15} bimetallic materials. The acetylene conversion ($X^{measured}_{C_2H_2}$) measured at $T_{oven} = 150°C$ at different t_{OS}, shown in Fig. 5A and B, indicates full conversion of acetylene and is consistent with the analysis of the SG rules in Fig. 4A, which indicates that the activity for these systems only drops at very low amounts of palladium. The measured and predicted ethylene selectivity ($S^{measured}_{C_2H_4}$ and $S^{SISSO}_{C_2H_4}$, respectively) are shown in Fig. 5A and B as blue and red squares. Note that the predictions are slightly different from the predictions shown in Fig. 4B, since they were made using the values of D_μ (6.03 and 7.04, respectively) and w_{metal} (0.0153 and 0.0161) measured by TEM and EDX-SEM, respectively, for the synthesized Pd_1Ag_{12} and Pd_1Ag_{15} materials. Pd_1Ag_{12} and Pd_1Ag_{15} present an initial measured selectivity of 67 and 68 %, respectively, at 27 min. These selectivities slightly drop to 61% at 405 min. These values are comparable to the selectivity of the material Pd_1Ag_9 (65 % at 405 min), which was the material with the best performance among those used to train the model. The SISSO model predicts the ethylene selectivity for Pd_1Ag_{12} and Pd_1Ag_{15} with good accuracy. This is to be expected, since the composition of these materials are only slightly different from the Pd_1Ag_9 material, which was included in the training data. However, the Pd_1Ag_{15} material is predicted to have a slightly higher selectivity compared to the actually measured one. The SISSO model is also correct in that no significant change in selectivity occurs with t_{OS}. Only minor decreases in selectivity are observed in experiment.

In addition to the bimetallic systems, we also used the AI models to predict the reactivity of the Pd-Ag-Cu system as an example of a trimetallic material. This system is represented as an orange triangle in Fig. 4A. The SG constraints indicate that high $X_{\mathrm{C_2H_2}}$ can be achieved for most of the compositions, with the exception of silver- and copper-rich systems. The predicted $S^{\mathrm{SISSO}}_{\mathrm{C_2H_4}}$ values (Fig. 4D and E) indicate that the selectivity increases with t_{OS}, in particular for silver- and copper-rich compositions. The composition for which the selectivity presents its maximum value at $t_{\mathrm{OS}} = 400$ min is approximately $Pd_1Ag_5Cu_1$. The trimetallic $Pd_1Ag_5Cu_1$ system was experimentally synthesized, characterized by TEM and EDX-SEM and tested. The measured D_μ and w_{metal} for $Pd_1Ag_5Cu_1$ is equal to 6.03 nm and 0.0145, respectively. It presents an initial measured selectivity of 62 % at 27 min, which slightly increases to 65% at 405 min (Fig. 5C). Thus, the trimetallic $Pd_1Ag_5Cu_1$ material is as selective as the most selective material of the training set (Pd_1Ag_9). Even though the predicted selectivity values deviate significantly from the measurements at the beginning of the reaction, the SISSO model correctly predicts the ethylene selectivity of $Pd_1Ag_5Cu_1$ at long t_{OS}, e.g., $t_{\mathrm{OS}} > 300$ min. Besides, the model is qualitatively correct in that the selectivity increases with t_{OS}. These results indicate the potential of the AI approach to guide the design of materials.

We note that the quality of the predictions of SISSO models can be improved with more data acquired in a systematic way using active learning.[55] By training ensembles of SISSO models with different subsets of training data or with different subsets of primary features, a statistical distribution of analytical expressions can be created. These expressions all describe the training data well. However, the predictions of the different expressions for new materials present a variance, which quantifies the uncertainty of the predictions. This uncertainty can be used to steer the acquisition of new data in unexplored regions of the materials space. Thus, larger portions of materials space can be systematically covered. Furthermore, the active-learning framework can also focus the acquisition of new data corresponding to the high-performance region of the materials-property map to improve the reliability of the SISSO description in this most important region.

IV. CONCLUSION

We developed and applied a focused AI approach to model the time-on-stream evolution of the measured reactivity of palladium-based alloys synthesized mechanochemically and applied in the selective hydrogenation of equally concentrated acetylene-ethylene streams. Out of 21 measured and calculated candidate descriptive parameters reflecting the reaction conditions and materials' mesoscopic, bulk, and (sub)surface properties, we identified key physical parameters correlated with the reactivity. The key materials' parameters identified by AI highlight the crucial influence of surface and subsurface carbon and hydrogen chemistry on the selectivity, among other underlying processes. Guided by the AI models, new bimetallic and trimetallic alloys were designed and tested experimentally. The materials proposed by AI and tested experimentally present ethylene selectivity comparable to the best performing catalysts in the training set. Besides, the AI models predict the evolution of their reactivity with time on stream with good accuracy. Further investigations center currently on the mechanochemical synthesis of immiscible Pd-Ru alloys which are predicted by the AI models to outperform the rather typical Pd-Ag systems in this reaction. In fact, mechanochemistry might enable the synthesis of immiscible systems, e.g., Pd-Pt and Pt-Au.[35] The focused data-centric approach introduced in this paper allows to navigate across immense pools of possible catalytic materials while taking into account their time-on-stream behavior, making it a valuable tool in catalyst and process design.

ACKNOWLEDGMENTS

We acknowledge the support from the Max Planck-Cardiff Centre on the Fundamentals of Heterogeneous Catalysis (FUNCAT). J. M. M. is thankful for a Kekulé scholarship by Fonds der Chemischen Industrie (FCI). L. F. and M. S. acknowledge funding from the NOMAD Center of Excellence (European Union's Horizon 2020 research and innovation program, Grant Agreement No. 951786) and the ERC Advanced Grant TEC1p (European Research Council, Grant Agreement No 740233). We gratefully acknowledge Eko Budiyanto, Alexander Hopf, Adrian Schlüter and Norbert Pfänder (MPI CEC)

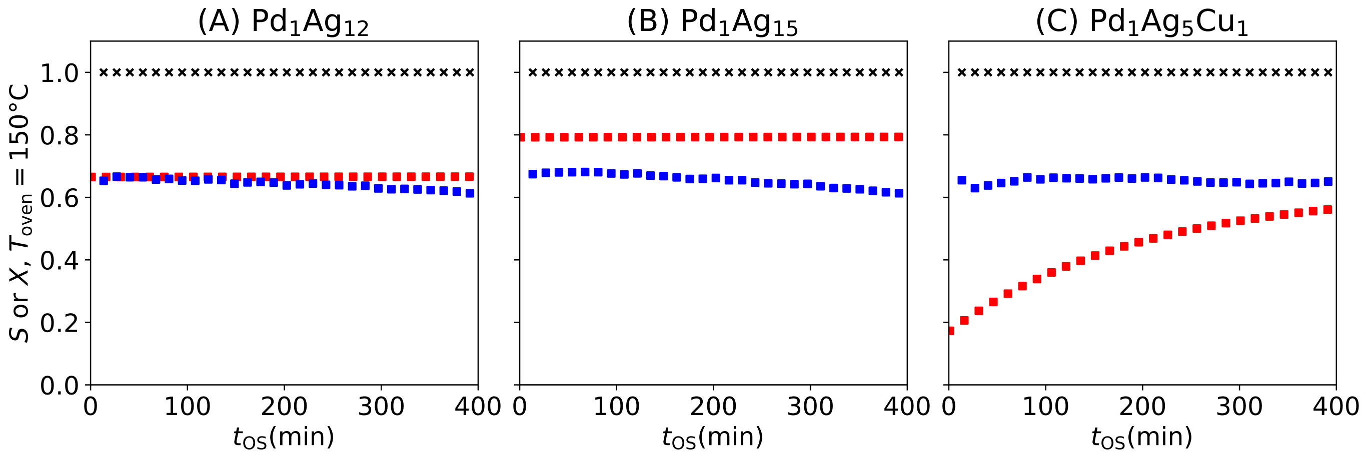

FIG. 5. The reactivity of two new bimetallic materials (A) $\mathrm{Pd_1Ag_{12}}$ and (B) $\mathrm{Pd_1Ag_{15}}$ and one new trimetallic material (C) $\mathrm{Pd_1Ag_5Cu_1}$ was measured. The measured ethylene selectivity $S^{\rm measured}_{\rm C_2H_4}$ (blue squares) is compared with the predictions of the SISSO model $S^{\rm SISSO}_{\rm C_2H_4}$ (red squared). The measurements and predictions are made for $T_{\rm oven} = 150°\mathrm{C}$. The black crosses indicate $X^{\rm measured}_{\rm C_2H_2}$.

for TEM, Silvia Palm for SEM, Jan Ternieden, Sebastian Leiting and Claudia Weidenthaler for XRD and XPS. We thank Henning Buchholz and Sabrina Brunner for assistance in the lab. We are grateful to the fine mechanic workshop (led by Dirk Ullner), the technical laboratories and central safety (led by Niels Theyssen) and the chromatography and electrophoresis service department (led by Philipp Schulze) for the support in the maintenance of the setup and its sophisticated safety structure.

*foppa@fhi-berlin.mpg.de

---

[1] Gabor A. Somorjai and Jeong Y. Park. Molecular factors of catalytic selectivity. *Angewandte Chemie International Edition*, 47(48):9212–9228, 2008.

[2] John Meurig Thomas. Heterogeneous catalysis: Enigmas, illusions, challenges, realities, and emergent strategies of design. *The Journal of Chemical Physics*, 128(18):182502, 05 2008.

[3] Hans-Joachim Freund, Gerard Meijer, Matthias Scheffler, Robert Schlögl, and Martin Wolf. Co oxidation as a prototypical reaction for heterogeneous processes. *Angewandte Chemie International Edition*, 50(43):10064–10094, 2011.

[4] Robert Schlögl. Heterogeneous catalysis. *Angewandte Chemie International Edition*, 54(11):3465–3520, 2015.

[5] Annette Trunschke, Giulia Bellini, Maxime Boniface, Spencer J Carey, Jinhu Dong, Ezgi Erdem, Lucas Foppa, Wiebke Frandsen, Michael Geske, Luca M Ghiringhelli, et al. Towards experimental handbooks in catalysis. *Topics in Catalysis*, 63:1683–1699, 2020.

[6] Lucas Foppa, Luca M. Ghiringhelli, Frank Girgsdies, Maike Hashagen, Pierre Kube, Michael Hävecker, Spencer J. Carey, Andrey Tarasov, Peter Kraus, Frank Rosowski, Robert Schlögl, Annette Trunschke, and Matthias Scheffler. Materials genes of heterogeneous catalysis from clean experiments and artificial intelligence. *MRS Bulletin*, 46:1016–1026, Nov 2021.

[7] Lucas Foppa, Christopher Sutton, Luca M. Ghiringhelli, Sandip De, Patricia Löser, Stephan A. Schunk, Ansgar Schäfer, and Matthias Scheffler. Learning design rules for selective oxidation catalysts from high-throughput experimentation and artificial intelligence. *ACS Catalysis*, 12:2223, 2022.

[8] Lucas Foppa, Frederik Ruther, Michael Geske, Gregor Koch, Frank Girgsdies, Pierre Kube, Spencer J Carey, Michael Havecker, Olaf Timpe, Andrey V Tarasov, et al. Data-centric heterogeneous catalysis: identifying rules and materials genes of alkane selective oxidation. *Journal of the American Chemical Society*, 145(6):3427–3442, 2023.

[9] Giulia Bellini, Gregor Koch, Frank Girgsdies, Jinhu Dong, Spencer J. Carey, Olaf Timpe, Gudrun Auffermann, Matthias Scheffler, Robert Schlögl, Lucas Foppa, and Annette Trunschke. Co oxidation catalyzed by perovskites: The role of crystallographic distortions highlighted by systematic experiments and artificial intelligence. *Angewandte Chemie International Edition*, n/a(n/a):e202417812.

[10] Ray Miyazaki, Kendra S. Belthle, Harun Tüysüz, Lucas Foppa, and Matthias Scheffler. Materials genes of co2 hydrogenation on supported cobalt catalysts: An artificial intelligence approach integrating theoretical and experimental data. *Journal of the American Chemical Society*, 146(8):5433–5444, 2024.

[11] Christopher Sutton, Mario Boley, Luca M. Ghiringhelli, Matthias Rupp, Jilles Vreeken, and Matthias Scheffler. Identifying domains of applicability of machine learning models for materials science. *Nature Communications*, 11(1):4428, Sep 2020.

[12] Stefan Wrobel. An algorithm for multi-relational discovery of subgroups. In *European Conference on Principles*

of Data Mining and Knowledge Discovery, 1997.

[13] Martin Atzmueller. Subgroup discovery. *WIREs Data Mining and Knowledge Discovery*, 5(1):35–49, 2015.

[14] Bryan R Goldsmith, Mario Boley, Jilles Vreeken, Matthias Scheffler, and Luca M Ghiringhelli. Uncovering structure-property relationships of materials by subgroup discovery. *New Journal of Physics*, 19(1):013031, jan 2017.

[15] Lucas Foppa and Luca M. Ghiringhelli. Identifying outstanding transition-metal-alloy heterogeneous catalysts for the oxygen reduction and evolution reactions via subgroup discovery. *Topics in Catalysis*, 65:196, Jan 2022.

[16] Lucas Foppa and Matthias Scheffler. Coherent collections of rules describing exceptional materials identified with a multi-objective optimization of subgroups. *arXiv:2403.18437*, 2024.

[17] Herzain I. Rivera-Arrieta and Lucas Foppa. Rules describing co2 activation on single-atom alloys from dft-meta-gga calculations and artificial intelligence. *ACS Catalysis*, page 10.1021/acscatal.4c07178, 2025.

[18] Runhai Ouyang, Stefano Curtarolo, Emre Ahmetcik, Matthias Scheffler, and Luca M. Ghiringhelli. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Physical Review Materials*, 2(8):083802, 2018.

[19] Runhai Ouyang, Emre Ahmetcik, Christian Carbogno, Matthias Scheffler, and Luca M Ghiringhelli. Simultaneous learning of several materials properties from incomplete databases with multi-task sisso. *Journal of Physics: Materials*, 2(2):024002, mar 2019.

[20] Lucas Foppa, Thomas A. R. Purcell, Sergey V. Levchenko, Matthias Scheffler, and Luca M. Ghiringhelli. Hierarchical symbolic regression for identifying key physical parameters correlated with bulk properties of perovskites. *Physical Review Letters*, 129:055301, Jul 2022.

[21] Matthias Scheffler, Christian Carbogno, Luca M Ghiringhelli, et al. Sisso++: A c++ implementation of the sure-independence screening and sparsifying operator approach. *Journal of Open Source Software*, 7(71):3960, 2022.

[22] Thomas A. R. Purcell, Matthias Scheffler, and Luca M. Ghiringhelli. Recent advances in the SISSO method and their implementation in the SISSO++ code. *The Journal of Chemical Physics*, 159(11):114110, 09 2023.

[23] Ioan-Teodor Trotuş, Tobias Zimmermann, and Ferdi Schüth. Catalytic reactions of acetylene: A feedstock for the chemical industry revisited. *Chemical Reviews*, 114(3):1761–1782, Feb 2014.

[24] Michael Bowker, Serena DeBeer, Nicholas F. Dummer, Graham J. Hutchings, Matthias Scheffler, Ferdi Schüth, Stuart H. Taylor, and Harun Tüysüz. Advancing critical chemical processes for a sustainable future: Challenges for industry and the max planck–cardiff centre on the fundamentals of heterogeneous catalysis (funcat). *Angewandte Chemie International Edition*, 61(50):e202209016, 2022.

[25] Evangelos Delikonstantis, Elorri Igos, Michael Augustinus, Enrico Benetto, and Georgios D. Stefanidis. Life cycle assessment of plasma-assisted ethylene production from rich-in-methane gas streams. *Sustainable Energy Fuels*, 4:1351–1362, 2020.

[26] Alan J McCue and James A Anderson. Recent advances in selective acetylene hydrogenation using palladium containing catalysts. *Frontiers of Chemical Science and Engineering*, 9:142–153, 2015.

[27] Madelyn R. Ball, Keishla R. Rivera-Dones, Elise B. Gilcher, Samantha F. Ausman, Cole W. Hullfish, Edgard A. Lebrón, and James A. Dumesic. Agpd and cupd catalysts for selective hydrogenation of acetylene. *ACS Catalysis*, 10(15):8567–8581, Aug 2020.

[28] Kefeng Xie, Kai Xu, Mingqiang Liu, Xiaohong Song, Shengyuan Xu, and Huayan Si. Catalysts for selective hydrogenation of acetylene: a review. *Materials Today Catalysis*, page 100029, 2023.

[29] Hannah Schreyer, Rene Eckert, Sarah Immohr, Jacopo deBellis, Michael Felderhoff, and Ferdi Schüth. Milling down to nanometers: A general process for the direct dry synthesis of supported metal catalysts. *Angewandte Chemie International Edition*, 58(33):11262–11265, 2019.

[30] Jacopo De Bellis, Michael Felderhoff, and Ferdi Schüth. Mechanochemical synthesis of supported bimetallic catalysts. *Chemistry of Materials*, 33(6):2037–2045, 2021.

[31] Amol P. Amrute, Jacopo De Bellis, Michael Felderhoff, and Ferdi Schüth. Mechanochemical synthesis of catalytic materials. *Chemistry – A European Journal*, 27(23):6819–6847, 2021.

[32] Klara Sophia Kley, Jacopo De Bellis, and Ferdi Schüth. Selective hydrogenation of highly concentrated acetylene streams over mechanochemically synthesized pdag supported catalysts. *Catalysis Science and Technology*, 13:119–131, 2023.

[33] Karen J. Ardila-Fierro and José G. Hernández. Sustainability assessment of mechanochemistry by using the twelve principles of green chemistry. *ChemSusChem*, 14(10):2145–2162, 2021.

[34] Amol P. Amrute, Zbigniew Łodziana, Hannah Schreyer, Claudia Weidenthaler, and Ferdi Schüth. High-surface-area corundum by mechanochemically induced phase transformation of boehmite. *Science*, 366(6464):485–489, 2019.

[35] Jacopo De Bellis, Hilke Petersen, Jan Ternieden, Norbert Pfänder, Claudia Weidenthaler, and Ferdi Schüth. Direct dry synthesis of supported bimetallic catalysts: A study on comminution and alloying of metal nanoparticles. *Angewandte Chemie International Edition*, 61(40):e202208016, 2022.

[36] I.-T. Trotus. *PhD thesis*. PhD thesis, Ruhr-Universität Bochum, 2016.

[37] Ioan-Teodor Trotuş, Tobias Zimmermann, Nicolas Duyckaerts, Jan Geboers, and Ferdi Schüth. Butadiene from acetylene–ethylene cross-metathesis. *Chemical Communications*, 51:7124–7127, 2015.

[38] Özgül Agbaba, Ioan-Teodor Trotuş, Wolfgang Schmidt, and Ferdi Schüth. Light olefins from acetylene under pressurized conditions. *Industrial & Engineering Chemistry Research*, 62(4):1819–1825, Feb 2023.

[39] Jonathan M. Mauß and Ferdi Schüth. On the role of anions in solid catalysts with ionic liquid layer (scill) for the selective hydrogenation of highly concentrated acetylene streams. *ChemSusChem*, n/a(n/a):e202401593.

[40] Bjørk Hammer and Jens Kehlet Nørskov. Theoretical surface science and catalysis—calculations and concepts. *Advances in catalysis*, 45:71–129, 2000.

[41] Detre Teschner, János Borsodi, Attila Wootsch, Zsolt Révay, Michael Hävecker, Axel Knop-Gericke, S. David Jackson, and Robert Schlögl. The roles of subsurface carbon and hydrogen in palladium-catalyzed alkyne hydrogenation. *Science*, 320(5872):86–89, 2008.

[42] Felix Studt, Frank Abild-Pedersen, Thomas Bligaard, Rasmus Z. Sørensen, Claus H. Christensen, and Jens K. Nørskov. On the role of surface modifications of palladium catalysts in the selective hydrogenation of acetylene. *Angewandte Chemie International Edition*, 47(48):9299–9302, 2008.

[43] Bo Yang, Robbie Burch, Christopher Hardacre, Gareth Headdock, and P. Hu. Influence of surface structures, subsurface carbon and hydrogen, and surface alloying on the activity and selectivity of acetylene hydrogenation on pd surfaces: A density functional theory study. *Journal of Catalysis*, 305:264–276, 2013.

[44] Jeff Greeley and Manos Mavrikakis. Surface and subsurface hydrogen: Adsorption properties on transition metals and near-surface alloys. *The Journal of Physical Chemistry B*, 109:3460–3471, 2005.

[45] Philippe Sautet and Fabrizio Cinquini. Surface of metallic catalysts under a pressure of hydrocarbon molecules: Metal or carbide? *ChemCatChem*, 2(6):636–639, 2010.

[46] E. J. Candes and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[47] Lance J. Nelson, Gus L. W. Hart, Fei Zhou, and Vidvuds Ozoliņš. Compressive sensing as a paradigm for building physics models. *Physical Review B*, 87(3):035125, 2013.

[48] C. van Heerden. Autothermic processes. *Industrial & Engineering Chemistry*, 45(6):1242–1247, Jun 1953.

[49] C. van Heerden. The character of the stationary state of exothermic processes. *Chemical Engineering Science*, 8(1):133–145, 1958.

[50] Emanuele Vignola, Stephan N. Steinmann, Ahmad Al Farra, Bart D. Vandegehuchte, Daniel Curulla, and Philippe Sautet. Evaluating the risk of c–c bond formation during selective hydrogenation of acetylene on palladium. *ACS Catalysis*, 8(3):1662–1671, Mar 2018.

[51] Leilei Zhang, Maoxiang Zhou, Aiqin Wang, and Tao Zhang. Selective hydrogenation over supported metal catalysts: From nanoparticles to single atoms. *Chemical Reviews*, 120(2):683–733, Jan 2020.

[52] Wei Huang, William Pyrz, Raul F. Lobo, and Jingguang G. Chen. Selective hydrogenation of acetylene in the presence of ethylene on k+--zeolite supported pd and pdag catalysts. *Applied Catalysis A: General*, 333(2):254–263, 2007.

[53] Lichen Liu and Avelino Corma. Metal catalysts for heterogeneous catalysis: From single atoms to nanoclusters and nanoparticles. *Chemical Reviews*, 118(10):4981–5079, May 2018.

[54] SN Tripathi, SR Bharadwaj, and SR Dharwadkar. The pd-ru system (palladium-ruthenium). *Journal of phase equilibria*, 14:638–642, 1993.

[55] Akhil S. Nair, Lucas Foppa, and Matthias Scheffler. Materials-discovery workflows guided by symbolic regression: Identifying acid-stable oxides for electrocatalysis. *arXiv:2412.05947*, 2024.