

Catalysis-Hub.org: An Open Electronic Structure Database for Surface Reactions

Kirsten T. Winther^{†,‡*}, Max J. Hoffmann^{†,‡},
Osman Mamun^{†,‡}, Jacob R. Boes^{†,‡}, Jens K. Nørskov^{†,‡,§},
Michal Bajdich[†] and Thomas Bligaard^{†*}

November 15, 2018

[†] SUNCAT Center for Interface Science and Catalysis, SLAC, National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California, 94025, United States

[‡] SUNCAT Center for Interface Science and Catalysis, Department of Chemical Engineering, Stanford University, Stanford, California 94305, United States

[§] Department of Physics, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark

*corresponding authors: Kirsten T. Winther (winther@stanford.edu) and Thomas Bligaard (bligaard@stanford.edu)

Abstract

We present a new open repository for chemical reactions on catalytic surfaces, available at <https://www.catalysis-hub.org>. The featured Surface Reactions database contains more than 100,000 chemical adsorption and reaction energies obtained from electronic structure calculations, and is continuously being updated with new datasets. In addition to providing quantum-mechanical results for a broad range of reactions and surfaces from different publications, the database features a systematic, large-scale study of chemical adsorption and hydrogenation on bimetallic alloy surfaces. The database contains reaction specific information, such as the surface composition and reaction energy for each reaction, as well as the atomic surface geometries used in the calculations together with the calculation parameters and output, which are essential for data reproducibility. Data can be accessed from the web-interface as well as from a Python API providing direct access from a local workstation. This enables researchers to efficiently use the data as a basis for further calculations and to generate surrogate models for accelerating the discovery of catalytic materials for sustainable energy applications.

Introduction

Electronic structure methods based on density functional theory (DFT) hold the promise to enable a deeper understanding of reaction mechanisms and reac-

tivity trends for surface catalyzed chemical and electrochemical processes and eventually to accelerate discovery of new catalysts.

As computational resources continue to increase, together with improved access to large-scale supercomputer resources, the rate of electronic structure calculations performed will also increase, resulting in an accelerated rate of data generation [1]. This leads us to a new paradigm of data-driven computational catalysis research where the increasing amount of computational data can be utilized to train surrogate models to direct and accelerate efforts for the identification of improved catalysts. Through collaborative efforts and the development of open-source databases and software tools the prospect of catalysis knowledge engines for automated catalyst design and discovery is realistic [2].

In the regime of data-driven research, it is important that data can be accessed efficiently and selectively so that meaningful subsets can be leveraged to make new computational insights into catalyst design. Therefore, development of advanced approaches for storing and accessing relevant data are critical. One of the crucial steps towards this goal is the establishment of curated open access databases [3]. Several open access databases for electronic structure calculations have emerged in the last decade with great success, such as the Inorganic Crystal Structure Database (ICSD), Materials Project [4], Open Quantum Materials Database (OQMD) [5], the Novel Materials Discovery (NoMaD) repository [6], Automatic Flow for materials discovery (AFLOW) [7], the ioChem-BD platform [8] and the Computational Materials Repository (CMR) [9, 10, 11]. While the databases mentioned above primarily feature calculations for bulk structures, 2D materials and gas phase molecules, the representation of specialized properties such as catalytic activity introduces additional complexity. A proper representation requires a surface specific database, where reaction energies, chemical species, surface facets, and surface compositions have been parsed, by tying together the output of several calculations.

A specialized database for chemical reactions on surfaces was previously achieved by CatApp [12], where reaction and activation energies for more than 3,000 reactions on primarily closed-packed transition metal surfaces are accessible from a web browser. However, CatApp does not store the atomic structures or the detailed computational settings and output of the electronic structure calculations which are important for data reproducibility. Additionally, the atomic structures are essential for constructing high-quality models of catalytic activity since the catalytic properties of a surface are determined by the local atomic structure of the active site for a reaction.

Here, we present a specialized database for reaction and activation energies for chemical reactions on catalytic surfaces which includes electronic structure geometries and contains more than 100,000 adsorption and reaction energies. The database is available from the web platform Catalysis-Hub (<https://www.catalysis-hub.org>) that serves as a framework for sharing data within the catalysis community and computational tools for theoretical catalysis research. The platform features several other applications (apps) for plotting results, creating and analyzing catalytic structure calculations, such as setting up new surface and adsorbate geometries [13] and making machine learning

a

Surface Reactions

3 entries

Reactants \rightarrow Products \times Surface \times Facet

CH3CHO* CH3CHO* Rh

Matching Reactions (3)

| Geometry | Reaction | Reaction Energy | Activation Energy | Surface | Facet | Sites | XC Functional |
|----------|---|-----------------|-------------------|---------|-------|-------|-----------------------|
| | CH3CHOH* + * \rightarrow CH3CHO* + H* | -0.10 eV | 0.83 eV | Rh | 111 | | BEEF-vdW/ASE-Espresso |
| | CH3CHO(g) + * \rightarrow CH3CHO* | -0.89 eV | | Rh | 211 | | BEEF-vdW/ASE-Espresso |
| | CH3CHO(g) + * \rightarrow CH3CHO* | -0.41 eV | | Rh | 111 | | BEEF-vdW/ASE-Espresso |

10* 1-3 of 3 < >

b

Reaction Geometries

Rh36C2H4O Rh36C2H5O Rh36H Rh36C2H5O Rh36

CH3CHO @ Rh36 [111]

Figure 1: Web interface to the Surface Reactions database. a) Searching for reactions by choosing reactants, products, surface composition and/or facet. b) Visualizing atomic geometries for a selected reaction.

predictions for adsorption energies [14]. A full description of the platform is beyond the scope of this work which will primarily focus on the Surface Reactions database.

Results

In this section we will cover the functionality of the Surface Reactions database and give an overview of the database infrastructure and types of data stored. Also, different approaches for accessing data will be given. The last part of the section will describe the featured datasets, with focus on a recent study on surfaces of bimetallic alloys.

The Surface Reactions database

The purpose of the Surface Reactions database is to store adsorption, reaction and reaction barrier energies for processes occurring on catalytic surfaces obtained from electronic structure calculations. It also serves to make these results easily available to the public and other researchers for aiding in new catalytic discoveries. We are focusing on chemical reactions of interest for sustainable energy applications such as conversion of CO₂ and synthetic gas to fuels [15, 16], electrochemical fuel cells [17, 18], and production of fuels and chemicals from electrochemical approaches [19]. The catalytic materials of interest for these applications includes transition metals and alloys, metal-oxides and oxy-hydroxides, perovskites, layered 2D materials, and metal-chalcogenites.

In order to model heterogeneous catalytic systems from electronic structure theory researchers often turn to simplified surface slab structures to approximate catalyst surfaces. Here, the different adsorption and active sites are sampled in order to generalize the model to more realistic conditions, such as catalytic nanoparticles [20]. The calculation of a reaction energy typically involves at least three electronic structure calculations, including the clean surface slab, the surface with the adsorbed species, and gas phase references of the adsorbate. Also, prior to calculating adsorption energies, the structure of the surface slab should be optimized starting from a bulk calculation. Furthermore, additional calculations are necessary in order to obtain the transition state geometry that determines the activation barrier for a reaction. We handle this complexity by storing a collection of processed reaction and activation energies which are linked to all the atomic geometries involved, including the bulk structure if available.

In the Surface Reactions app at, <https://www.catalysis-hub.org/energies>, the user can search for chemical reactions on surfaces by specifying reactants, products, surface composition, and/or surface facet. The result of the search will be returned as a list of rows in the browser showing the surface composition, the chemical equation of the reaction, reaction energy, activation energy (when present), and user specified adsorption sites. Selecting the geometry symbol to the left of a given row will expand the result, allowing users to browse the atomic structures linked to the reaction and see publication info and calculation details.

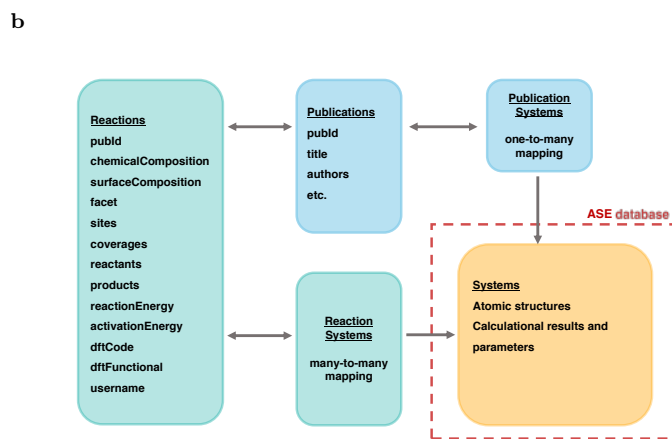
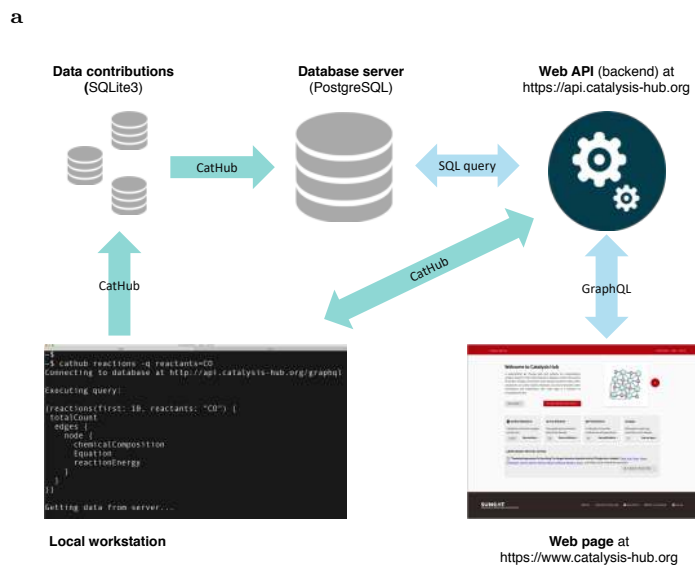


Figure 2: a) Schematic overview of the database platform, showing the relation between the database server, backend and frontend applications. b) Database table layout.

An example of a reaction search is given in Figure 1, showing the results for reactions on Rhodium containing CH₃CHO* on the product side. The atomic structures can be spatially repeated in the browser for better visualization and downloaded in a large set of formats including cif, JSON, xyz, VASP POSCAR, CASTEP Cell, Quantum Espresso input, and other. Additionally, reaction energies and structures can be browsed publication-wise by using the Publications button accessible from the front page.

Database infrastructure

An overview of the database infrastructure is shown in Figure 2. The platform consists of a database server where the data is stored, a web application programming interface (API) that handles queries to the database, and a frontend application which serves the main web page. Data fetching from the backend is handled by a graph based query language, GraphQL [21], and the Catalysis Hub (CatHub) software package provides an additional Python API, making the data accessible from a local workstation. Data is stored in a relational database instance (PostgreSQL) on an Amazon Web Server where it is continuously backed up. In structured query language (SQL), data is stored in a collection of ordered tables, where selections on properties (columns) can be applied to return a subset of rows and columns from the tables. A schematic overview of the SQL tables used for the Surface Reactions database is shown in Figure 2. Separate tables are used to store publications, reactions, and atomic structures (systems), allowing for one-to-many and many-to-many mappings between these properties. The Reactions table contains reaction specific info, so that fast queries on chemical composition of the surface, reaction energy, and adsorption sites can be performed. Each reaction is linked to the atomic structures involved (such as adsorbed species, empty slabs, gas phase references, and bulk structure) in the systems table. Also, both reactions and atomic structures are linked to the corresponding entry in the publication table.

The atomic structures are stored in ASE database layout. ASE[22] is a popular software package for setting up and managing atomic structures, with interfaces to a large set of popular electronic structure codes. The ASE database is developed specifically for storing atomic structures, computational results and parameters, making it a natural choice for handling the atomic structures of reaction intermediates.

The Systems table of the ASE database contains information about the geometry (such as atomic numbers, positions, and constraints), calculator settings, and the output of the calculation (such as energy, forces, and magnetic moments). The detailed table layout and data types are provided in Table 2 and 1 in the Methods section.

Data accessibility

In addition to the primary web interface, data can also be accessed directly from the web API at <https://api.catalysis-hub.org/graphql>, which uses

the graph based query language GraphQL. Here, queries to the database can be typed directly into the browser. A detailed description of the database backend is provided in the methods section.

Furthermore, the CatHub software package provides a convenient interface to the database. Acting as a Python API, CatHub be used for data fetching and manipulation directly from a Python script. In practice, this is done by sending a GraphQL query to the API as a HTTP request, which will return a JSON dictionary with the selected data. A code snippet with an example of how to access reaction energies directly from a Python script is shown below,

```
from cathub.query import get_reactions
reactions = get_reactions(n_results=10,
                          chemicalComposition='~Ni',
                          reactants='CO2'
                          write_db=True)
```

which will return a JSON dictionary with the data for the first ten reactions involving carbon dioxide on the reactants side, on surfaces containing Nickel. The `write_db=True` option will copy the selected data to a local database file, including all the atomic structures used in the reactions that can be examined with the CatHub module and the ASE database. In addition, the CatHub Command Line Interface (CLI) provides access directly from the terminal.

A wrapper around the ASE database CLI allows users to access the atomic structures in the database directly from their local terminal. The query below will select all atomic structures from the database containing both Silver and Strontium without any restriction on stoichiometric ratio,

```
cathub ase --args AgSr --gui
```

where the `-gui` option will open the selected atomic structures directly in the ASE GUI visualizer.

Featured dataset

The Surface reactions database currently contains results from more than 46 publications and datasets, all available at <https://www.catalysis-hub.org/publications>, where reactions from publication can be browsed with visualization of atomic geometries. Most of the datasets stem from already published work and contain a direct link to the publication homepage via the digital object identifier (DOI). Also, a collection of soon to be published/just submitted datasets are made available. Recently featured datasets include studies of syngas to C+ Oxygenates conversion on transition metals [16], oxygen reduction and hydrogen oxidation on metal-doped 2D materials [18] and a study of solvated protons at the water-metal interface [23]. As a whole, the database contains roughly 700 different chemical reactions, involving more than 150 chemical species and 3,500 different catalytic material surfaces.

In order to facilitate data exchange and promote publications of the catalysis and surface science communities the database is open to contributions of

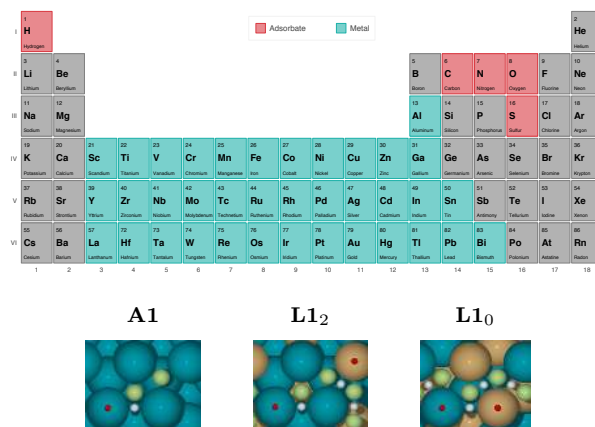


Figure 3: The periodic table outlining 5 adsorbate elements and the 37 metals included in the bimetallic alloys dataset, including six metals from group 13-15, seventeen transition metals, and Lanthanum. Surface geometry and enumerated adsorption sites for the three structures are provided in the lower panel, where top, bridge, and hollow sites are shown in red, white, and green, respectively.

first-principles reaction calculations from the community under <https://www.catalysis-hub.org/upload>. Contributions will appear in the Publications section with its own permanent URL featuring authorship and may contain a link to a concomitant publication. The data upload process is handled by the CatHub module, which aides in organizing a given folder of output files into a structure suitable for uploading. Any self-contained dataset (gas phase references, empty slab, adsorbate geometry) of ASE readable DFT output files are welcome.

Large scale study of bimetallic alloys

The largest dataset features a new high-throughput study of chemical adsorption and hydrogenation on more than 2000 bimetallic alloy and pure metal surfaces. This dataset is continuously expanded and updated in a collaborative effort in the SUNCAT Group* and is available online at <https://www.catalysis-hub.org/publications/WinterOpen2018>. The data can be accessed with the tools described in the previous section.

The metallic alloys were constructed by combining 37 selected metals and transition metals, shown in Fig 3, in the L₁₂ and L₁₀ Strukturbericht designation, which corresponds to face-centered cubic crystal structures with A₃B and AB stoichiometries, respectively. The 37 pure metals in the A1 (fcc) structure were also included in addition to the 1,998 bimetallic alloys resulting from all

*The dataset was generated by Mamun Osman, Jacob R. Boes, and Kirsten Winther at SUNCAT, SLAC National Accelerator Laboratory

possible combinations. The metal surfaces were modeled by cleaving three-layer slabs with a (111) termination for A1 and L1₂ and a (101) termination for L1₀, although this termination is also referred to as the (111) miller index when cleaved from the cubic bulk unit cell which is not the standard conventional form [24, 25].

Atomic adsorption of H, C, N, O, and S were studied for all 2,035 metal surfaces. The number of unique adsorption sites are 4, 9, and 10 for the A₁, L1₂, and L1₀ surfaces respectively, shown in Fig 3. This gives a total of 96,015 unique surfaces, adsorbate and site combinations (including the empty slabs), where roughly 65,000 calculations are completed so far. Also, the adsorption of the hydrogenated species CH, NH, CH₂, NH₂, SH, OH and H₂O has been studied for a smaller subset of alloy surfaces, where alloys formed from 16 metals of particular interest for catalysis has been chosen, with approximately 20,000 calculations completed. The CatGen [13] module of CatKit [26] was used to generate the slab structures from optimized bulk systems and to systematically enumerate all high-symmetry adsorption sites.

Examples of calculated adsorption energies are given in Fig. 4, showing adsorption energies for most stable sites for atomic carbon, oxygen, and nitrogen on the L1₀ and L1₂ alloys. In Fig. 4(a) and (b) the adsorption energies are plotted as a function of metal A and B, that are arranged on an improved Pettifor scale, [27, 28] with small adjustments for magnetic elements, which ensures a smooth variation of the adsorption energies with composition. The sampled surfaces cover an extensive range in adsorption energies, spanning more than 5 eV with strong (low values) adsorption for early transition metal alloys (top left corner) and weak adsorption (high values) for noble and late metals (lower right corner). In the case of L1₀ alloys, the AB and BA structures are identical, such that the plot in Fig. 4(a) is symmetric across the diagonal. For the L1₂ alloys, the A₃B composition gives rise to a larger dependence of the A-metal, seen as smaller horizontal variation of the adsorption energies in Fig. 4(b). We note that the preferred adsorption sites also show a large variation across the metal alloy space. Regions with preference towards top, bridge, and three-fold hollow sites can be identified.

Another approach for visualizing adsorption energy trends is to plot the adsorption energy of two adsorbates against each other, which often gives rise to linear scaling relationships for similar surface geometries. Utilizing scaling relationships is a well established approach in theoretical catalysis to model and understand catalytic activity and selectivity [29, 30]. In the lower panel of Fig. 4 the correlation between the adsorption of carbon with (c) oxygen and (d) nitrogen is shown. Metals containing a partially filled *d*-band versus a filled or empty *d*-band is labeled as *d*- and *non-d* metals respectively. All alloys containing a *non-d* metal are labeled as *non-d* alloys. While a close to linear relationship between the adsorption of C and O can be seen for the *d* and *non-d* pure metals separately, the correlation between the atomic adsorption energies on the alloys are more complicated, emphasizing the need for more sophisticated methods for modelling these systems, such as data-driven approaches. We also note that traditional scaling relations are intended only for the most stable site

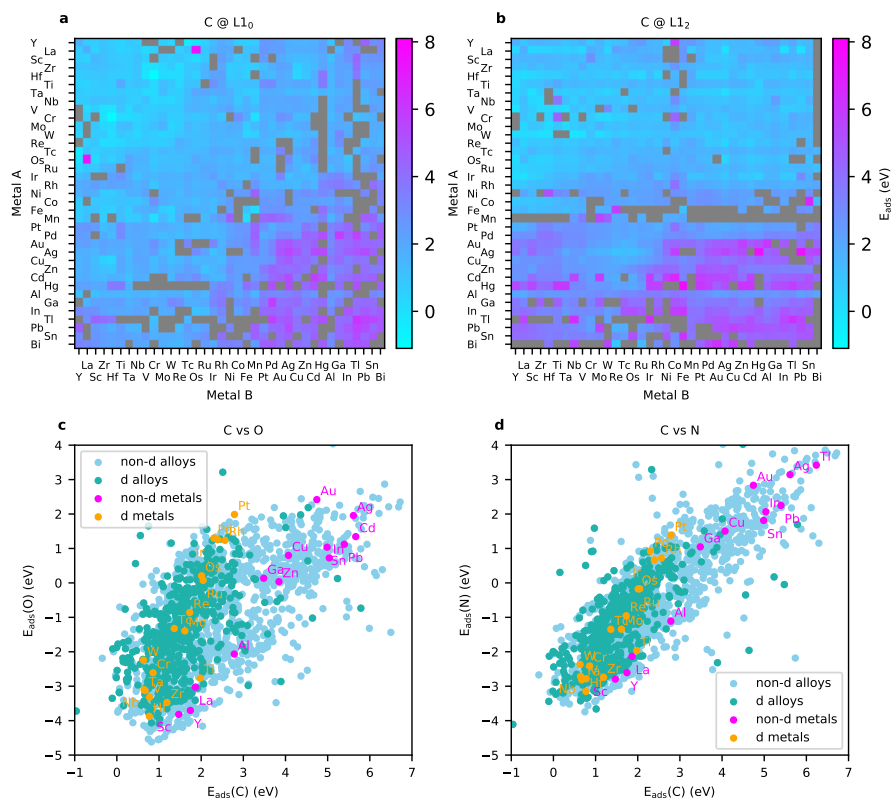


Figure 4: Adsorption energies of atomic carbon for: a) the 666 L_{10} (A_2B) and b) the 1332 L_{12} (A_3B) alloy surfaces. Results for the 37 pure metals are shown along the diagonal. Adsorption energies of atomic (c) oxygen and (d) nitrogen are plotted with respect to the C adsorption energy for all materials. References are taken with respect to the reactions: $\text{CH}_4(\text{g}) - 2\text{H}_2(\text{g}) + * \rightarrow \text{C}^*$, $0.5\text{N}_2(\text{g}) + * \rightarrow \text{N}^*$ and $\text{H}_2\text{O}_2(\text{g}) - \text{H}_2(\text{g}) + * \rightarrow \text{O}^*$ with all species adsorbed to their lowest energy sites.

of a given surface while these plots include all kinetically trapped minima in energy. The script used to plot Fig. 4(c) by fetching the data directly with the CatHub module is provided in the methods section.

Discussion

We believe that the Surface Reactions database will be of great benefit to the scientific community and will aid researchers in their search for new materials for catalysis and sustainable energy applications. By creating a platform for sharing recent scientific results we are enabling the community to efficiently build on top of each other’s work with direct access to the computational data from several channels. To these ends, community contributions are strongly encouraged.

We wish to ensure that the database has both substantial breadth as well as depth; i.e. covering a large range of different materials and reactions with a large collection of results for a particular material and reaction. An increased diversity of data is accomplished by featuring data from a large number of publications. This is demonstrated through the many small and focused datasets which have already been uploaded. This also ensures that the database contains catalytic materials from recent cutting-edge research which will be further facilitated by contributions from a diversity of research groups. Through the data upload functionality in the CatHub software package this contributing process is greatly simplified. On the other hand, the generation of surrogate modes, such as machine learning algorithms, generally require vast amount of systematic generated data. Therefore, database also contains large computationally-consistent datasets targeted for machine learning purposes, such as the bimetallic alloys dataset described in this work.

A concern regarding the breadth and depth of data is how to obtain reliable reaction energy barriers for a large set of reactions and materials. Since the energy barriers determine the kinetics (or reaction rate) of a chemical reaction, a good prediction is important for getting a quantitative measure for the catalytic activity and selectivity. Due to the high computational cost of determining the transition state of energy barriers, only a small fraction of the reactions have an associated energy barrier calculated from DFT. Therefore, our focus has been on populating the database with a large set of adsorption energies, which are significantly cheaper to compute and can serve as descriptors to model reaction energies and barriers[31]. In time, advanced machine learning techniques and targeted kinetic systems of interest will supply more accurate barrier energetics to the existing data.

Methods

All computational results available from the Surface Reactions database have been obtained from electronic structure calculations, using DFT. The detailed

calculator settings vary between datasets.

For the bimetallic alloys dataset, calculations were performed in the Quantum Espresso (QE) electronic structure code [32], using the BEEF-vdw exchange correlation functional [33], a 500 eV plane-wave cutoff, and a 5,000 eV density cutoff. Monkhorst-Pack k -point grids of (12, 12, 12) for bulk and (6, 6, 1) for slab calculations were used, together with 0.15 eV Fermi smearing. Spin-polarized calculations were performed only for alloys containing Fe, Ni, Co, and Mn, with an initial guess for the atomic magnetic moments close to the atomic magnetization of the pure metal in the lowest close-packed (fcc, bcc, or hcp) structure. For A_1 and $L1_2$, lattice constants were obtained from bulk alloy calculations with an equation of state combined with an energy minimization in QE; however, for $L1_0$ structures we used a variable cell optimization in QE with a high plane wave cutoff (800 eV) and then used the resulting lattice constants as initial guess for the final energy minimization with respect to lattice constant parameters - i.e., 'a' and 'c' - using the Scipy fmin optimizer [34]. Slab geometries were optimized by fixing the two bottom layers and allowing the top layer and adsorbates to relax. Due to the large number of calculations, job submissions were handled with FireWorks [35] and the CatFlow submodule of CatKit, that provides a FireWorks interface to all electronic structures calculators supported by ASE.

Upon relaxation we found that reconstructions of the metal surfaces, e.g. horizontal sliding or dissociation of the top layer from the slab, are quite common. Also, we found that the adsorbates often reorient into sites which differed from the initial sites. The relaxed geometries were therefore post-processed with a tailored classification method to label reconstructed surfaces and reclassify the adsorption sites.

Only non-reconstructed surfaces have been used to generate adsorption energies, although, as the reconstructed structures are still of interest for model generation, the atomic structures are still included within the database.

Data structure

The full layout of the SQL tables used for the Surface Reactions database is given in Table 1 and Table 2, listing the full set of columns for the ASE database systems table, and the Surface Reactions database specific tables, respectively. The ASE database was updated in connection to this project in order to utilize native array and JSONB datatypes for PostgreSQL, where the JSONB datatype is a binary JSON format that stores user-defined keys and values in a search-optimized way. This enables faster queries on user defined key-value-pairs. The ARRAY data type is used to store arrays such as the atomic positions and numbers, which ensures that selections on the chemical composition (and potentially local atomic structure in the vicinity of adsorbates) can be executed directly in SQL.

| column name | data type |
|----------------------|------------------------|
| id | integer |
| uniqueId | text |
| ctime | double precision |
| mtime | double precision |
| username | text |
| numbers | integer[] |
| positions | double precision[][] |
| cell | double precision[][] |
| pbcs | integer |
| initialMagmoms | double precision[] |
| initialCharges | double precision[] |
| masses | double precision[] |
| tags | integer[] |
| momenta | double precision[] |
| constraints | text |
| calculator | text |
| calculatorParameters | jsonb |
| energy | double precision |
| freeEnergy | double precision |
| forces | double precision[][] |
| stress | double precision[] |
| dipole | double precision[] |
| magmoms | double precision[] |
| magmom | double precision |
| charges | double precision[] |
| keyValuePairs | jsonb |
| data | jsonb |
| natoms | integer |
| fmax | double precision |
| smax | double precision |
| volume | double precision |
| mass | double precision |
| charge | double precision |

Table 1: PostgreSQL table structure of the systems table of the ASE database, listing column names and datatypes. Array datatypes are marked with “[]” for a 1D array and “[][]” for a 2D array. The JSONB datatype saves dictionaries in a binary format that is fast to process, and allows for fast queries on key value pairs.

| table name | column name | data type |
|-------------------|---------------------|-----------|
| reactions | id | integer |
| | chemicalComposition | text |
| | surfaceComposition | text |
| | facet | text |
| | sites | jsonb |
| | coverages | jsonb |
| | reactants | jsonb |
| | products | jsonb |
| | reactionEnergy | numeric |
| | activationEnergy | numeric |
| | dftCode | text |
| | dftFunctional | text |
| | username | text |
| | pubId | text |
| textsearch | tsvector | |
| reactionSystems | name | text |
| | energyCorrection | numeric |
| | aseId | text |
| | id | integer |
| publications | id | integer |
| | pubId | text |
| | title | text |
| | authors | jsonb |
| | journal | text |
| | volume | text |
| | number | text |
| | pages | text |
| | year | smallint |
| | publisher | text |
| | doi | text |
| | tags | jsonb |
| | pubtextsearch | tsvector |
| publicationSystem | aseId | text |
| | pubId | text |

Table 2: SQL table structure for the Surface Reactions database specific tables.

```

1- {
2-   reactions(first: 3, products: "CH3CO",
3-     order: "activationEnergy") {
4-     edges {
5-       node {
6-         Equation
7-         chemicalComposition
8-         reactionEnergy
9-       }
10-    }
11-  }
12- }
13-
{
  "data": {
    "reactions": {
      "edges": [
        {
          "node": {
            "Equation": "CH3CHO* + * -> CH3CO* + H**",
            "chemicalComposition": "Rh36",
            "reactionEnergy": -0.586442627391
          }
        },
        {
          "node": {
            "Equation": "CH3CHO* + * -> CH3CO* + H**",
            "chemicalComposition": "Ir36",
            "reactionEnergy": -0.685817585196
          }
        },
        {
          "node": {
            "Equation": "CH3CHO* + * -> CH3CO* + H**",
            "chemicalComposition": "Pt36",
            "reactionEnergy": -0.832361438188
          }
        }
      ]
    }
  }
}

```

Figure 5: Example of a GraphQL query for reactions, executed in the web interface at <https://api.catalysis-hub.org>.

Frontend and backend applications

The main web page is served as a frontend application that runs on a Node.js instance on the Heroku Cloud Application Platform [36]. The frontend source code is implemented using the React framework, and is available open-source on GitHub [37]. Atomic structures are visualized in the browser using the ChemDoodle[38] web component.

Retrieval of data from the database server is managed by a backend application, which acts as a web API. In practice, the backend is a collection of software that runs on a Python framework Heroku. The backend is build with Flask [39], a microframework for web development in Python, and uses a Python SQL toolkit, SQLAlchemy[40], for connecting to the database server and handling relations (such as foreign key constraints and many-to-many mappings) between SQL tables.

Data fetching from the backend to the frontend is handled with GraphQL[21], which is a graph based query language developed by Facebook as an alternative to representational state transfer (REST). It provides simple and user friendly data-fetching, where the request is sent as a string in JSON-like format that specifies the data to be selected. A JSON object with the same data structure as the request is returned. The backend can be accessed at <https://api.catalysis-hub.org/graphql>, where GraphQL queries can be typed directly into the browser. An example of such a query is given in Figure 5, where the first three reactions involving CH_3CO on the right hand side (in order of increasing activation energy) is returned. The structure of the output is the same as the input, and a

join with the publication table is easily performed by adding the publications field inside reactions. This conveniently enables access to the publication info for a particular reaction.

As described in the main text, data from the Surface Reactions database can also be accessed directly from a Python script by using the CatHub module. The code used for plotting the data shown in Figure 4, is made available as a tutorial from https://github.com/SUNCAT-Center/CatHub/tree/master/tutorials/1_bimetallic_alloys.

Acknowledgements

This work was supported by the U.S. Department of Energy, Chemical Sciences, Geosciences, and Biosciences (CSGB) Division of the Office of Basic Energy Sciences, via Grant DE-AC02-76SF00515 to the SUNCAT Center for Interface Science and Catalysis.

Competing financial interests

The authors declare no competing financial interests.

References

- [1] Robin Haunschuld, Andreas Barth, and Werner Marx. Evolution of DFT studies in view of a scientometric perspective. *Journal of cheminformatics*, 8(1):52, 2016.
- [2] Andrew J Medford, M Ross Kunz, Sarah M Ewing, Tammie Borders, and Rebecca Fushimi. Extracting knowledge from data through catalysis informatics. *ACS Catalysis*, 8(8):7403–7429, 2018.
- [3] Carles Bo, Feliu Maseras, and Núria López. The role of computational results databases in accelerating the discovery of catalysts. *Nature Catalysis*, page 1, 2018.
- [4] Anubhav Jain, Shyue P Ong, Geoffroy Hautier, Wei Chen, William D Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A Persson. The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013.
- [5] Scott Kirklin, James E Saal, Bryce Meredig, Alex Thompson, Jeff W Doak, Muratahan Aykol, Stephan Rühl, and Chris Wolverton. The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies. *npj Computational Materials*, 1:15010, 2015.

- [6] The NoMaD repository. <http://nomad-repository.eu/cms/>. The NoMaD (Novel Materials Discovery) repository was established to host, organize, and share materials data.
- [7] Stefano Curtarolo, Wahyu Setyawan, Gus L.W. Hart, Michal Jahnatek, Roman V. Chepulskii, Richard H. Taylor, Shidong Wang, Junkai Xue, Kesong Yang, Ohad Levy, Michael J. Mehl, Harold T. Stokes, Denis O. Demchenko, and Dane Morgan. Aflow: An automatic framework for high-throughput materials discovery. *Computational Materials Science*, 58(nil):218–226, 2012.
- [8] M Álvarez-Moreno, Cornelis De Graaf, Nuria Lopez, Feliu Maseras, Josep M Poblet, and Carles Bo. Managing the computational chemistry big data problem: the ioChem-BD platform. *Journal of chemical information and modeling*, 55(1):95–103, 2014.
- [9] Computational materials repository. <https://cmr.fysik.dtu.dk/>. This is an institutional repository for center for atomic-scale materials design (CAMD) at the Technical University of Denmark.
- [10] Sten Haastrup, Mikkel Strange, Mohnish Pandey, Thorsten Deilmann, Per S Schmidt, Nicki F Hinsche, Morten N Gjerding, Daniele Torelli, Peter M Larsen, Anders C Riis-Jensen, Jakob Gath, Karsten W Jacobsen, Jens Jørgen Mortensen, Thomas Olsen, and Kristian S Thygesen. The computational 2D materials database: high-throughput modeling and discovery of atomically thin crystals. *2D Materials*, 5(4):042002, 2018.
- [11] Per S Schmidt and Kristian S Thygesen. Benchmark database of transition metal surface and adsorption energies from many-body perturbation theory. *The Journal of Physical Chemistry C*, 122(8):4381–4390, 2018.
- [12] Jens S Hummelshøj, Frank Abild-Pedersen, Felix Studt, Thomas Bligaard, and Jens K Nørskov. CatApp: a web application for surface chemistry and heterogeneous catalysis. *Angewandte Chemie International Edition*, 51(1):272–274, 2012.
- [13] Jacob R. Boes, Osman Mamun, Kirsten T. Winther, and Thomas Bligaard. Graph theory approach to high-throughput surface adsorption structure generation. *ChemRxiv*, 2018.
- [14] Paul Jennings, Martin H Hansen, Jose A G Torres, and Thomas Bligaard. CatLearn: A machine learning environment for atomic-scale modeling in surface science and catalysis. <https://github.com/SUNCAT-Center/CatLearn>.
- [15] Velu Subramani and Santosh K Gangwal. A review of recent literature to search for an efficient catalytic process for the conversion of syngas to ethanol. *Energy & fuels*, 22(2):814–839, 2008.

- [16] Julia Schumann, Andrew J Medford, Jong S Yoo, Zhi-Jian Zhao, Pallavi Bothra, Ang Cao, Felix Studt, Frank Abild-Pedersen, and Jens K. Nørskov. Selectivity of synthesis gas conversion to C₂+ oxygenates on fcc(111) transition-metal surfaces. *ACS Catalysis*, 8(4):3447–3453, 2018.
- [17] Mark K Debe. Electrocatalyst approaches and challenges for automotive fuel cells. *Nature*, 486(7401):43, 2012.
- [18] Seoin Back, Ambarish R Kulkarni, and Samira Siahrostami. Single metal atoms anchored in two-dimensional materials: Bifunctional catalysts for fuel cell applications. *ChemCatChem*, 10(14):3034–3039, 2018.
- [19] Zhiyi Lu, Guangxu Chen, Yanbin Li, Haotian Wang, Jin Xie, Lei Liao, Chong Liu, Yayuan Liu, Tong Wu, Yuzhang Li, Alan C. Luntz, Michal Bajdich, and Yi Cui. Identifying the active surfaces of electrochemically tuned licoo₂ for oxygen evolution reaction. *Journal of the American Chemical Society*, 139(17):6270–6276, 2017.
- [20] Jens K Nørskov, Thomas Bligaard, Jan Rossmeisl, and Claus H Christensen. Towards the computational design of solid catalysts. *Nature chemistry*, 1(1):37, 2009.
- [21] GraphQL: A query language for your api. <https://graphql.org/>.
- [22] Ask H Larsen, Jens J Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, 2017.
- [23] Leanne D Chen, Michal Bajdich, J Mark P Martirez, Caroline M Krauter, Joseph A Gauthier, Emily A Carter, Alan C Luntz, Karen Chan, and Jens K Nørskov. Understanding the apparent fractional charge of protons in the aqueous electrochemical double layer. *Nature communications*, 9(1):3202, 2018.
- [24] Zhao-Xu Chen, Konstantin M Neyman, Kok H Lim, and Notker Rösch. CH₃O decomposition on PdZn(111), Pd(111), and Cu(111). a theoretical study. *Langmuir*, 20(19):8068–8077, 2004.
- [25] Alyssa J R Hensley, Sebastian Schneider, Yong Wang, and Jean-Sabin McEwen. Adsorption of aromatics on the (111) surface of PtM and PtM₃ (M = Fe, Ni) alloys. *RSC Advances*, 5(104):85705–85719, 2015.
- [26] CatKit: General purpose tools for high-throughput catalysis. <https://github.com/SUNCAT-Center/CatKit>.
- [27] David G Pettifor. A chemical scale for crystal-structure maps. *Solid State Communications*, 51(1):31–34, 1984.

- [28] Henning Glawe, Antonio Sanna, EKV Gross, and Miguel AL Marques. The optimal one dimensional periodic table: a modified pettifer chemical scale from data mining. *New Journal of Physics*, 18(9):093011, 2016.
- [29] Frank Abild-Pedersen, Jeff Greeley, Felix Studt, Jan Rossmeisl, TR Munter, Poul G Moses, Egill Skulason, Thomas Bligaard, and Jens K Nørskov. Scaling properties of adsorption energies for hydrogen-containing molecules on transition-metal surfaces. *Physical review letters*, 99(1):016105, 2007.
- [30] Andrew J Medford, Aleksandra Vojvodic, Jens S Hummelshøj, Johannes Voss, Frank Abild-Pedersen, Felix Studt, Thomas Bligaard, Anders Nilsson, and Jens K Nørskov. From the sabatier principle to a predictive theory of transition-metal heterogeneous catalysis. *Journal of Catalysis*, 328:36–42, 2015.
- [31] Jens K Nørskov, Thomas Bligaard, Ashildur Logadottir, Sune R Bahn, Lars B Hansen, Mikkel Bollinger, Hanne S Bengaard, Bjørk Hammer, Željko V Sljivancanin, Manos Mavrikakis, et al. Universality in heterogeneous catalysis. *Journal of Catalysis*, 209(2):275–278, 2002.
- [32] P Giannozzi, O Andreussi, T Brumme, O Bunau, M Buongiorno Nardelli, M Calandra, R Car, C Cavazzoni, D Ceresoli, M Cococcioni, N Colonna, I Carnimeo, A Dal Corso, S de Gironcoli, P Delugas, R A DiStasio Jr, A Ferretti, A Floris, G Fratesi, G Fugallo, R Gebauer, U Gerstmann, F Giustino, T Gorni, J Jia, M Kawamura, H-Y Ko, A Kokalj, E Küçükbenli, M Lazzeri, M Marsili, N Marzari, F Mauri, N L Nguyen, H-V Nguyen, A Otero de-la Roza, L Paulatto, S Poncé, D Rocca, R Sabatini, B Santra, M Schlipf, A P Seitsonen, A Smogunov, I Timrov, T Thonhauser, P Umari, N Vast, X Wu, and S Baroni. Advanced capabilities for materials modelling with quantum espresso. *Journal of Physics: Condensed Matter*, 29(46):465901, 2017.
- [33] Jess Wellendorff, Keld T. Lundgaard, Andreas Møgelhøj, Vivien Petzold, David D. Landis, Jens K. Nørskov, Thomas Bligaard, and Karsten W. Jacobsen. Density functionals for surface science: Exchange-correlation model development with bayesian error estimation. *Phys. Rev. B*, 85:235149, Jun 2012.
- [34] John A Nelder and Roger Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.
- [35] Anubhav Jain, Shyue P Ong, Wei Chen, Bharat Medasani, Xiaohui Qu, Michael Kocher, Miriam Brafman, Guido Petretto, Gian-Marco Rignanese, Geoffroy Hautier, Daniel Gunter, and Kristin A Persson. Fireworks: a dynamic workflow system designed for high-throughput applications. *Concurrency and Computation: Practice and Experience*, 27(17):5037–5059, 2015. CPE-14-0307.R2.
- [36] Heroku: Cloud application platform. <https://www.heroku.com/>.

- [37] SUNCAT Center for Interface Science and Catalysis. <https://github.com/SUNCAT-Center/>.
- [38] Melanie C. Burger. Chemdoodle web components: HTML5 toolkit for chemical graphics, interfaces, and informatics. *Journal of Cheminformatics*, 7(1):35, Jul 2015.
- [39] Armin Ronacher. Flask: Web development, one drop at the time. <http://flask.pocoo.org/>.
- [40] Michael Bayer. SQLAlchemy—the database toolkit for python. <https://www.sqlalchemy.org/>.