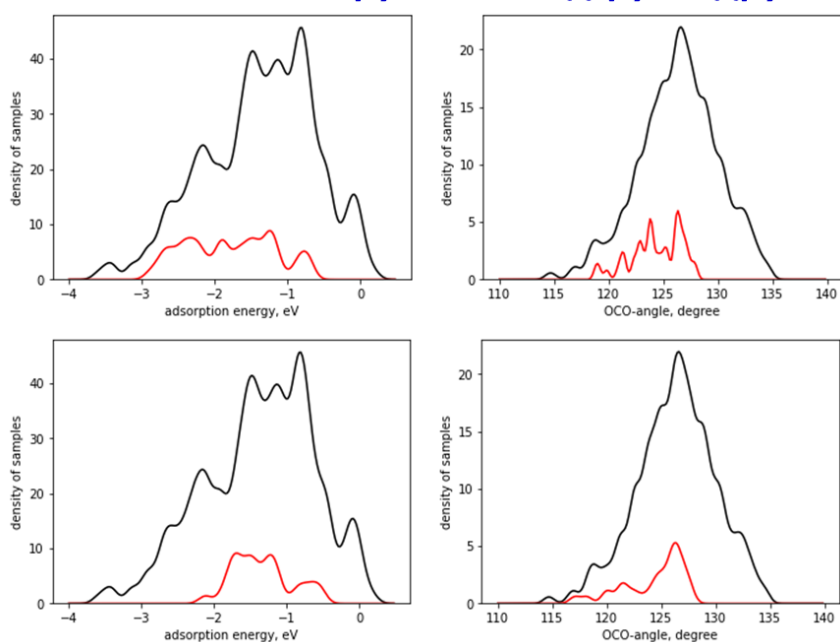


# ЛЕКЦИЯ 15. ПРЕДСКАЗАНИЕ ТЕРМОЭЛЕКТРИЧЕСКИХ МАТЕРИАЛОВ. ДЕСКРИПТОРЫ

Ранее мы обсуждали метод обнаружения подгрупп на примере поиска материалов для активации  $\text{CO}_2$  путем адсорбции на поверхностях. Было установлено, что если использовать как индикатор активации углы и не вводить никаких дополнительных условий, то мы находим подгруппу, в которой углы в среднем меньше, чем во всем наборе данных. При этом большая часть этих материалов характеризуется низкой энергией адсорбции  $\text{CO}_2$  на поверхностях. Это говорит о сильном связывании молекул с материалом, которое приводит к загрязнению поверхности и ее деактивации. Однако, если добавить условие, что энергия адсорбции для данной точки попадает в интервал, соответствующий принципу Сабатье, т.е. адсорбция должна быть не слишком сильной (качественная концепция химического гетерогенного катализа), то тогда находится подгруппа, в которой и энергии, и углы попадают в оптимальный интервал (рис. 1). Многие известные катализаторы, активирующие  $\text{CO}_2$ , принадлежат этой подгруппе.

## Обнаружение подгрупп: другой подход

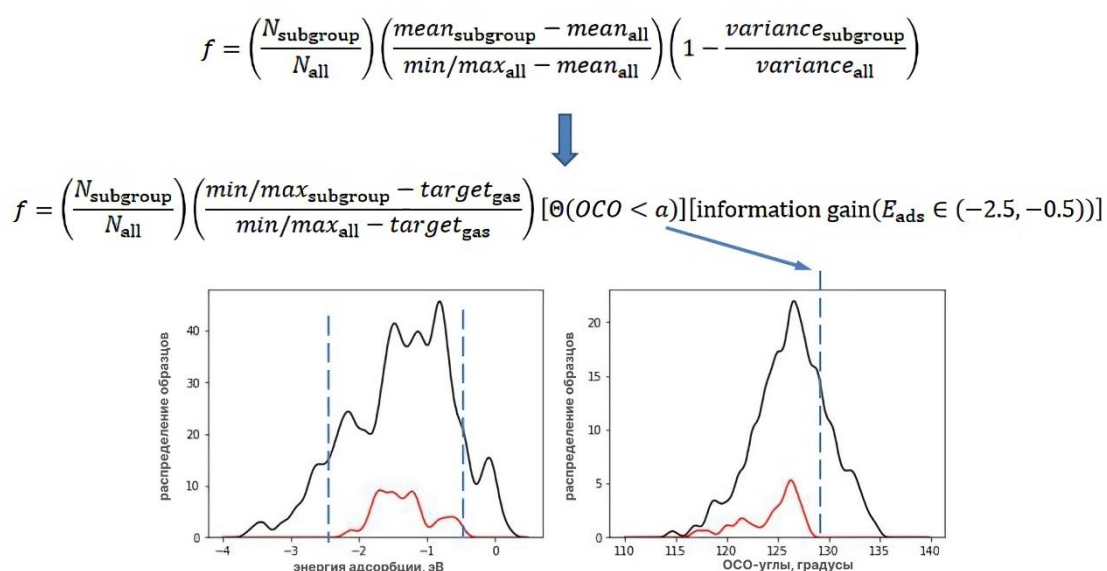


большой заряд Хиршфельда на поверхностном O, более низкая координация → меньшие углы

с ограничением энергии адсорбции: меньший заряд на поверхностном O, делокализованная электронная плотность, связывание O в  $\text{CO}_2$  с поверхностными катионами

Рисунок 1. Обнаружение подгрупп

Важное свойство метода обнаружения подгрупп: возможность легко подстроить функцию под наши нужды, то есть включить дополнительные условия. Данный метод дает несколько подгрупп одновременно. Например, в подгруппе с максимальной функцией полезности выделяет несколько других подгрупп, которые также являются статистически значимыми. Это может происходить по двум причинам. Во-первых, потому что есть основные признаки, которые используются для определения подгрупп, которые могут коррелировать друг с другом. Во-вторых, физически может быть несколько механизмов, которые обеспечивают, например, уменьшение угла ОСО при адсорбировании CO<sub>2</sub> (рис. 2).



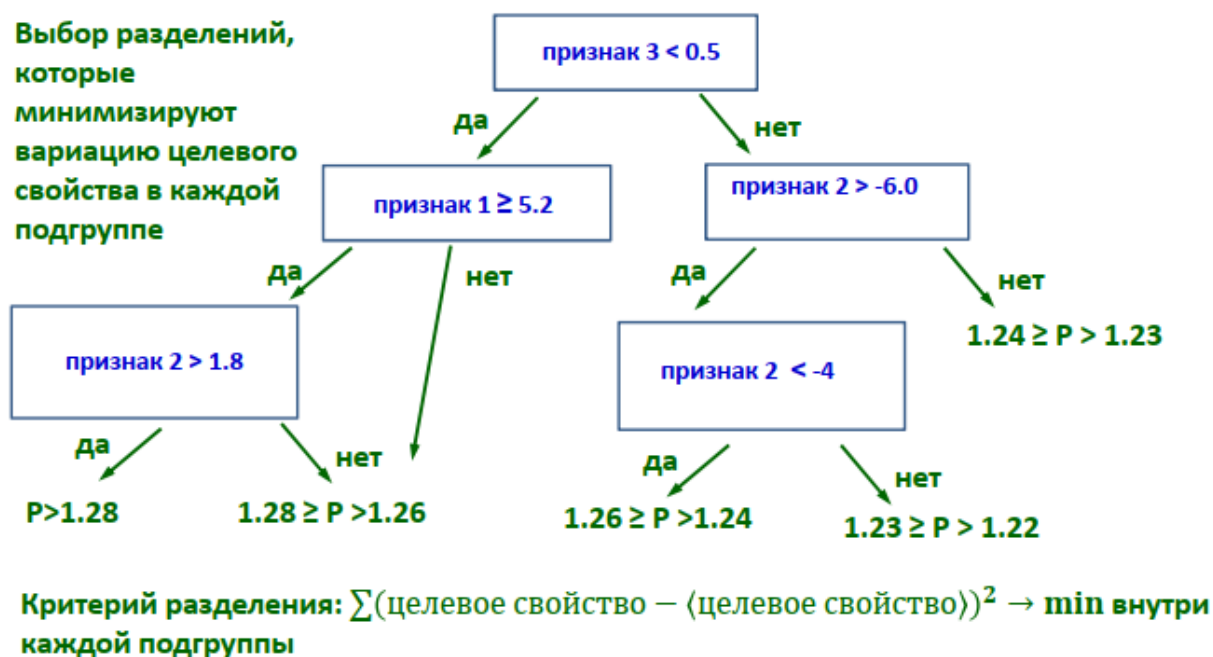
**Рисунок 2. Метод обнаружения подгрупп**

Возможность находить альтернативные пути оптимизации с помощью метода обнаружения подгрупп сильно отличает его от других аналогичных методов. Одним из таких методов является метод «деревьев решений» – стандартный метод машинного обучения, активно применяющийся в материаловедении.

Что же такое деревья решений? Например, у нас есть набор данных, и перед нами стоит задача найти хороший термоэлектрик. Сперва необходимо рассмотреть различные условия. С помощью критериев разделения, например

«да или нет», разобьем набор данных. Затем исследуем все признаки для каждой его части. Полученные последовательности называются ветвями, а конечные узлы-разделители - листьями - это дерево, которое растет вниз.

С помощью дерева решения можно также делать регрессию, при этом критерий разделения определяется как сумма по ошибкам в каждом разделении (целевое свойство для данной точки минус среднее целевое свойства в квадрате по подгруппе). Критерий разделения должен быть минимальным внутри каждой из подгрупп (рис. 3). Таким образом, весь набор данных разделяется на небольшие сетки, в каждом из которых можно предсказать примерное значение свойства.



*Рисунок 3. Пример дерева решений*

Перечислим основные свойства деревьев решений:

- простота понимания и интерпретации;
- глобальный подход (важное отличие от SGD, которое обнаруживает локально уникальные группы);
- легко перестараться с подгонкой (overfit) (можно использовать методы типа LASSO для решения этой проблемы);

– небольшое изменение данных может привести к большим изменениям в дереве;

– относительно неточно.

Это глобальный подход. Каждая из найденных подгрупп независима, поэтому вместе они минимизируют полную ошибку предсказания. Метод обнаружения подгрупп, в отличие от метода дерева решений, может дать альтернативные механизмы и локально уникальные группы.

К недостаткам деревьев решений относится возможность переопределение системы и высокая чувствительность - небольшое изменение данных может привести к большим изменениям в дереве – предсказание относительно неточное.

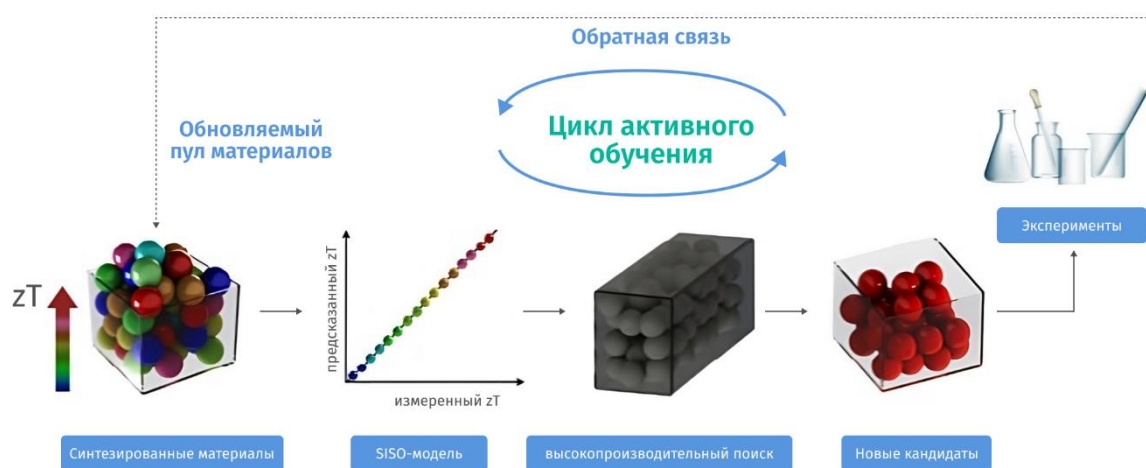
Чтобы улучшить точность метода и его предсказательную силу был предложен подход “случайный лес (Random forest)”, который состоит в том, что делается регрессия или классификация с помощью дерева решений для нескольких случайно выбранных подмножеств данных. Далее в каждом дереве при каждом разделении случайным образом выбирается фиксированное количество признаков, для которых определяется наилучшее разделение, и затем предсказания от полученных деревьев усредняются.

Почему “случайный лес” лучше? Это более точный подход, чем одно дерево («каждое дерево удерживает другие деревья от ошибок»), точность повышается, однако теряется интерпретируемость модели. Этот метод дает статистическую значимость и может использоваться для отбора первичных признаков для других подходов, таких как SISSO.

Интересные работы, в которых метод «случайного леса» использовался для определения наиболее важных структурных особенностей поверхности, определяющих ее стабильность, представлены на рисунке 4 в качестве примера.

**Рисунок 4. Примеры работ, использовавших метод “случайного леса”**

Существует другой подход для оптимизации термоэлектриков – путем анализа экспериментальных данных, схема которого представлена на рис. 5.



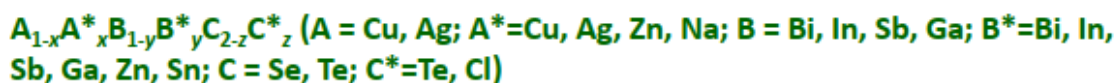
Сочетание экспериментов, анализа данных и высокопроизводительного скрининга

**Рисунок 5. Схема оптимизации термоэлектриков**

Она включает в себя следующие стадии: первоначально синтезируются материалы, затем, на основе полученных данных, проводится тренировка модели (SISO-Model), которая используется для высокопроизводительного поиска новых кандидатов с высоким показателем эффективности. Предсказанные материалы синтезируются, цикл замыкается. Так повторяется несколько раз. Данный цикл называется циклом активного обучения.

В одной из работ были исследованы поликристаллические халькогениды (термоэлектрики). Их общая формула представлена на рис. 6.

### Поликристаллические халькогениды:



A\*, B\*, C\*- примесные элементы

### *Рисунок 6. Поликристаллические халькогениды (термоэлектрики)*

Данные материалы получали методом вакуумного плавления – отжига в сочетании с искровым плазменным спеканием с использованием элементарных веществ (99,999%, Emei Semicon. Mater. Co., Ltd. Sichuan, CN) - это относительно простой синтез. Был исследован диапазон температур от 300 К до 870 К. Для данных веществ были проведены экспериментальные измерения коэффициентов Зеебека, электропроводности, теплопроводности, теплоемкости и плотности образцов различных составов при разных температурах. Это дало 600 точек данных, на основе которых было проведено три итерация активного обучения.

В качестве основных признаков использовались экспериментальные свойства атомов и элементарных твердых тел. Далее применяли стандартный метод SISO, то есть генерировали сложные признаки из первичных с помощью математических операторов, и затем искали путем сжатого зондирования наилучший дескриптор (рис. 7).

### SISSO (sure independence screening плюс sparsifying operator):

- 1) Экспериментальные свойства атомов и элементарных твердых тел (первичные признаки)
- 2) Сложные признаки генерируются из первичных с использованием математических операций {+, -, ·, /, log, exp, exp-, <sup>-1</sup>, <sup>2</sup>, <sup>3</sup>, √, <sup>3</sup>√, |-|} ~10<sup>10</sup> признаков
- 3) Поиск дескрипторов с помощью сжатого зондирования (скрининг определенной независимости с последующим выбором лучших одиночных, парных, тройных и т. д. комбинаций сложных признаков) :

$$P = c_0 + c_1 d_1 \quad P = c_0 + c_1 d_1 + c_2 d_2 \quad \dots \quad P = c_0 + \sum_{i=1}^n c_i d_i$$

$P - zT$

$d$  – сложные признаки - нелинейные комбинации первичных признаков

**10-кратная перекрестная проверка используется для определения оптимальной размерности дескриптора**

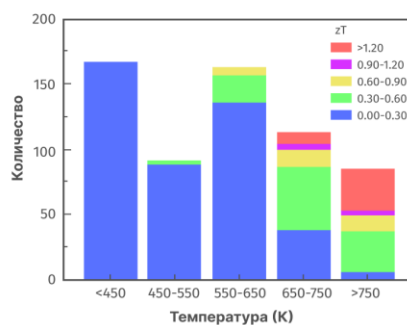
Ouyang, R. et al., Phys. Rev. Mater. 2, 083802 (2018)

### Рисунок 7. Схема поиска дескрипторов

Была использована десятикратная перекрестная проверка для определения оптимальной размерности. Первичные признаки представлены на рис. 8.

#### ПЕРВИЧНЫЕ ПРИЗНАКИ И ДАННЫЕ ДЛЯ $A_{1-x} A^*_x B_{1-y} B^*_y C_{2-z} C^*_z$

| Первичный признак            | Обозначение                                      |
|------------------------------|--|
| Температура (К)              | T  |
| Концентрация примесей        | $C_A, C_B, C_C(x, y, z) (0 < C_i < 0.5)$         |
| Электроотрицательность (эВ)  | $EN_A, EN_{A^*}, EN_B, EN_{B^*}, EN_C, EN_{C^*}$ |
| Энергия ионизации (эВ)       | $IE_A, IE_{A^*}, IE_B, IE_{B^*}, IE_C, IE_{C^*}$ |
| Теплота плавления (эВ)       | $HF_A, HF_{A^*}, HF_B, HF_{B^*}, HF_C, HF_{C^*}$ |
| Теплота парообразования (эВ) | $HV_A, HV_{A^*}, HV_B, HV_{B^*}, HV_C, HV_{C^*}$ |
| Атомный радиус (А)           | $AR_A, AR_{A^*}, AR_B, AR_{B^*}, AR_C, AR_{C^*}$ |
| Атомный вес (а.у.)           | $AW_A, AW_{A^*}, AW_B, AW_{B^*}, AW_C, AW_{C^*}$ |



Распределение данных для заключительной итерации активного обучения

Три итерация активного обучения, 602 итоговых точки (включая различные элементные составы, концентрации и температуры)

### Рисунок 8. Первичные признаки и данные для поликристаллических халькогенидов

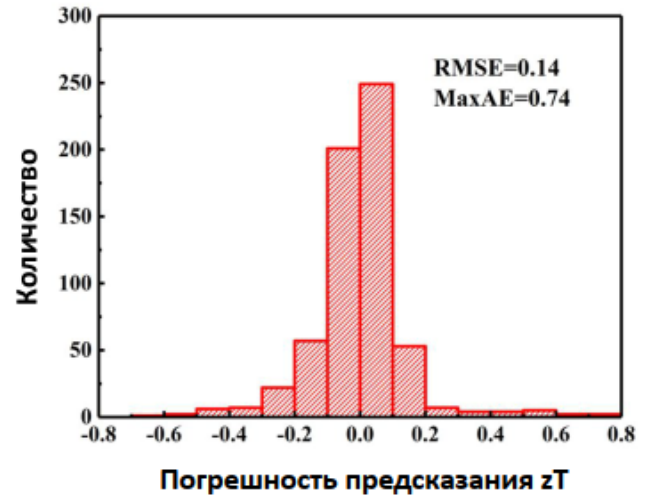
На рис. 8 представлено распределение данных в зависимости от температуры. При низких температурах отсутствуют материалы с высоким



показателем  $zT$ , с ростом температуры их число увеличивается (все материалы имеют одинаковую кристаллическую структуру).

На рис. 9 представлены компоненты дескриптора, показавшего лучшую точность при перекрестной проверки, а также представлена оценка важности данной компоненты.

| компонент дескриптора                           | оценка важности |
|---|-----------------|
| $+T^2/((1-C_{A^*}) \cdot AW_B)$                 | 0.65 (0.61)     |
| $+(T/AR_{B^*}) \cdot  AW_C - AW_{B^*} $         | 0.38 (0.17)     |
| $+((1-C_{A^*}) \cdot AR_B \cdot HV_A)/T$        | 0.21 (0.14)     |
| $- AW_B - AW_C  /  AW_A - AW_{B^*} $            | 0.18 (0.16)     |
| $-((1-C_{A^*}) \cdot AR_{A^*}) / (EN_B - EN_C)$ | 0.04 (0.06)     |
| $+(AW_A/T) /  HV_B - HV_{A^*} $                 | 0.06 (0.03)     |
| $-(T \cdot C_{B^*}) / (AR_A - AR_C)$            | 0.02 (0.01)     |



$$\text{оценка важности} = 1 - \frac{\text{RMSE (MaxAE) (все компоненты)}}{\text{RMSE (MaxAE) (все - 1 компоненты)}}$$

**Рисунок 9. Данные по компонентам дескриптора**

Чем выше этот показатель, тем более важной является компонента дескриптора. В скобках показана максимальная ошибка, без скобок - среднеквадратичная.

Интересна интерпретация этой модели (рис. 10). Например, в одном из компонентов дескриптора в знаменателе стоит  $(1-C_A)$  (единица минус концентрация примесей в подрешетке A), что может быть объяснено следующим образом: чем больше эта концентрация, тем выше беспорядок в этой подрешетке, тем ниже теплопроводность и выше показатель эффективности. Одновременно увеличение беспорядка вредно для электропроводности. При увеличении температуры происходит уменьшение электропроводности, что приводит к уменьшению показателя эффективности.



| компонент дескриптора                           | оценка важности |
|---|-----------------|
| $+T^2/((1-C_{A^*}) \cdot AW_B)$                 | 0.65 (0.61)     |
| $+(T/AR_{B^*}) \cdot  AW_C - AW_{B^*} $         | 0.38 (0.17)     |
| $+((1-C_{A^*}) \cdot AR_B \cdot HV_A)/T$        | 0.21 (0.14)     |
| $- AW_B - AW_C  /  AW_A - AW_{B^*} $            | 0.18 (0.16)     |
| $-((1-C_{A^*}) \cdot AR_{A^*}) / (EN_B - EN_C)$ | 0.04 (0.06)     |
| $+(AW_A/T) /  HV_B - HV_{A^*} $                 | 0.06 (0.03)     |
| $-(T \cdot C_{B^*}) / (AR_A - AR_C)$            | 0.02 (0.01)     |

$$zT = \frac{\alpha^2 \sigma T}{\kappa}$$

Беспорядок в А-подрешетке ( $C_{A^*} \uparrow \rightarrow \kappa \downarrow \rightarrow zT \uparrow$ )

но беспорядок вреден для электропроводности:  $C_{A^*}, T \uparrow \rightarrow \sigma \downarrow \rightarrow zT \downarrow$

Несоответствие размеров и веса между атомами В\* и А,В,С  $\rightarrow$  эффект дребезжания  $\uparrow, \kappa \downarrow, zT \uparrow$

$$\text{оценка важности} = 1 - \frac{\text{RMSE (MaxAE) (все компоненты)}}{\text{RMSE (MaxAE) (все - 1 компоненты)}}$$

*Рисунок 10. Интерпретация данных по компонентам дескриптора*

Также во многих компонентах дескриптора встречаются разности атомных весов или атомных радиусов элементов. Это позволяет учесть несоответствие размеров и весов между атомами (оно увеличивает эффект дребезжания), что приводит к уменьшению теплопроводности и увеличению показателя эффективности.

Модель SISO была применена для высокопроизводительного поиска. Первый цикл активного обучения был неуспешный, то есть предсказания были плохими, второй - несколько лучше, а третий - хороший. Всего было предсказано 10 тысяч значений показателя эффективности  $zT$  (рис. 11).

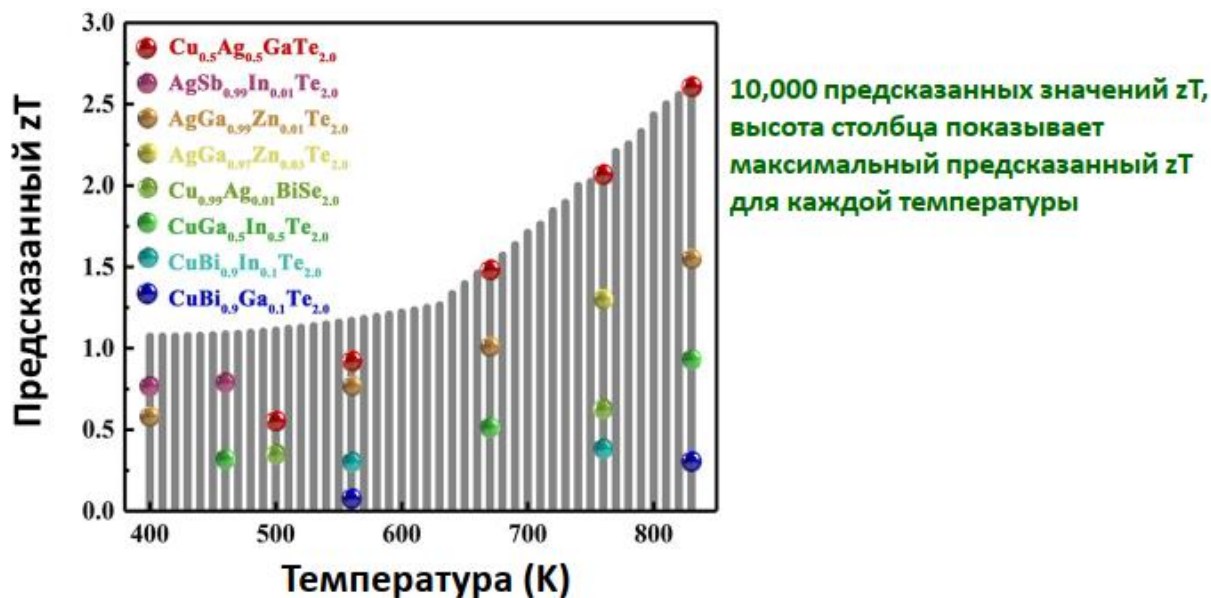


Рисунок 11. Результаты предсказаний с помощью модели SISSO

В итоге был получен показатель эффективности равный 2,8 zT – это очень высокое значение для легкосинтезируемых материалов (рис. 12).

Лучший предсказанный материал:  $\text{Cu}_{0.45}\text{Ag}_{0.55}\text{GaTe}_2$  (zT = 2.8 при 827 K)

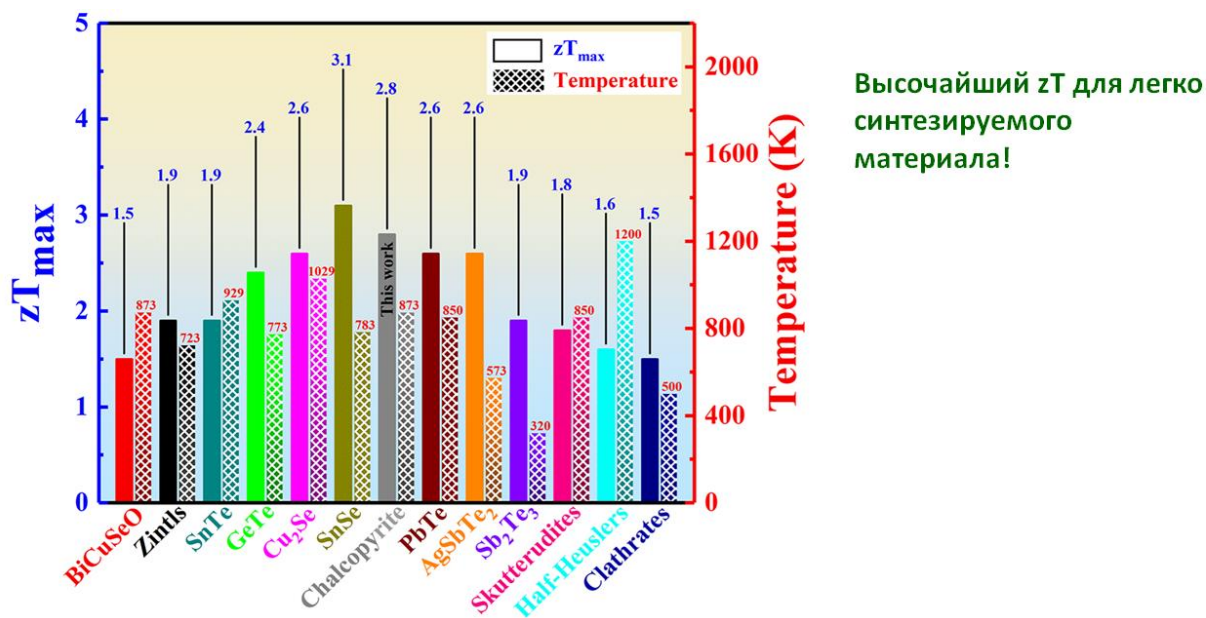
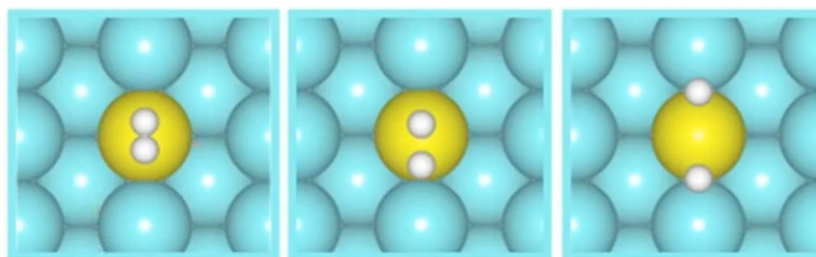


Рисунок 12. Значения показателей эффективности zT для различных материалов

Результаты были проверены независимо в нескольких лабораториях. Оценка  $zT_{\max}$  уменьшилась примерно до  $2zT$ , однако это тоже высокий показатель.

Далее рассмотрим, как еще можно использовать метод обнаружения подгрупп для интерпретации сложных моделей символьной регрессии на примере работы «Катализаторы на основе атомарно-дисперсных сплавов (SAACs)» (рис. 13).



**SAACs -- активные и селективные катализаторы гидрирования и других реакций**

**Баланс между диссоциацией  $H_2$  и связыванием H с поверхностью**

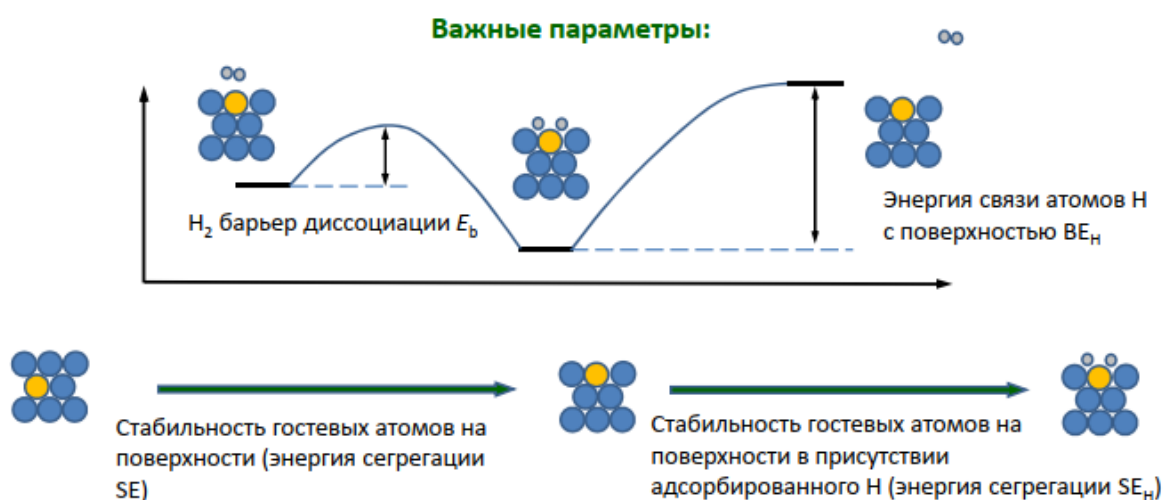
**Эффективное использование дорогих металлов**

Kyriakou, G. *et al.*, *Science* **335**, 1209-1212 (2012)

***Рисунок 13. Катализаторы на основе атомарно-дисперсных сплавов SAACs***

Атомарно-дисперсные сплавы - это сплавы, в которых один атом металла замещает другой атом металла или адсорбируется на его поверхности. Данные сплавы интересны своими уникальными свойствами. Они известны как активные и селективные катализаторы, например, процессов гидрирования. SAACs проявляют каталитические свойства за счет существования баланса между диссоциацией водорода  $H_2$  и связыванием атомов водорода H с поверхностью. Атомарно-дисперсные сплавы позволяют более эффективно использовать дорогие металлы.

Исследования SAACs было проведено с помощью искусственного интеллекта – сложность заключалась в многообразии поверхностных срезов, наличие различных основных и одиночных (диспергированных) элементов (рис. 14). Даже для бинарных сплавов у нас имеются десятки тысяч вариантов, для более сложных систем их больше.

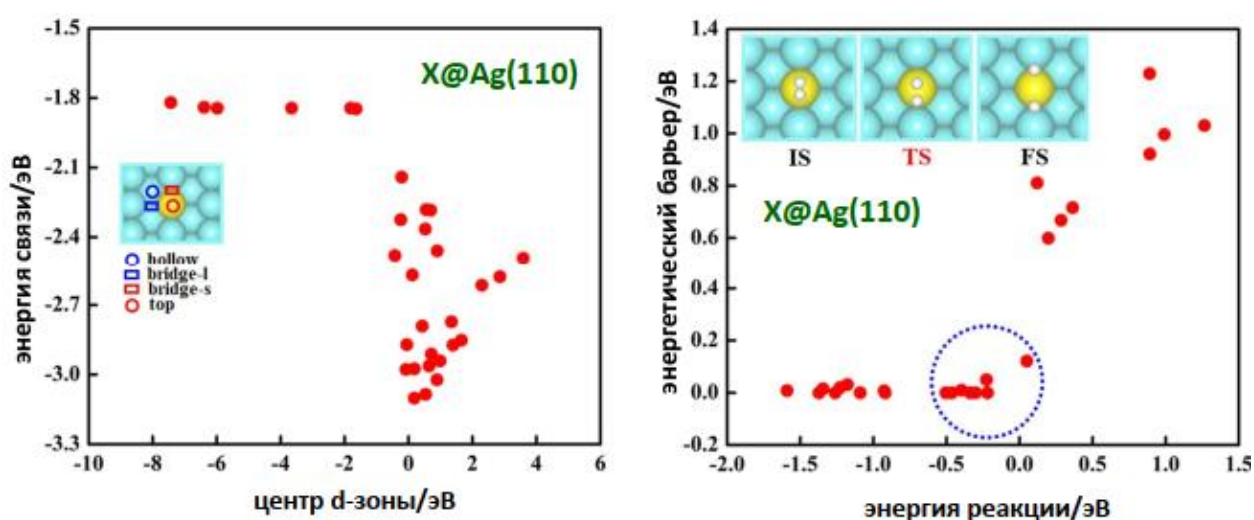


**Рисунок 14. Исследование атомарно-дисперсионных сплавов SAACs**

Мы хотим предсказать барьер диссоциации (должен быть как можно ниже для эффективной реакции) и энергию связи с поверхностью (должна удовлетворять принципу Сабатье). Важным показателем также является стабильность веществ - атомы должны оставаться на поверхности при катализе (не должны уходить вглубь вещества или слипаться друг с другом); стоит учитывать, что при катализе на поверхности присутствует водород, который может поменять стабильность материала.

Для чистых и простых поверхностей металлов наблюдается очень хорошая корреляция между энергией связи атома и центром d-зоны (хороший дескриптор). Однако такое простое изменение, как добавление одного единственного атома, меняет картину полностью (рис. 15, слева). Между энергией связи атома водорода и центром d-зоны обнаруживается плохая корреляция для различных диспергированных элементов на поверхности

серебра. Аналогична ситуация и для значений энергетического барьера – хорошей корреляции не наблюдается (рис. 15, справа).



**Рисунок 15. Корреляция между энергией связи атома/значением энергетического барьера и центром d-зоны**

Простые дескрипторы не отражают сложную физику катализаторов, необходимо использовать более сложные дескрипторы, основанные на признаках, представленных на рис. 16.

| класс                                     | название   | обозначение      |
|---|--|------------------|
| Атомные (основа [*] и примесь)            | Энергия НОМО и LUMO  | H*, L*, H, L     |
|   | Сродство к электрону (атомный радиус)                            | EA*(R*), EA(R)   |
|   | Потенциал ионизации  | IP*, IP          |
|   | Энергия связи H с одним атомом металла (атомов в димере металла) | EH*(EB*), EH(EB) |
|   | Длина связи H с одним атомом металла (в димере металла)          | dH*(dd*), dH(dd) |
| Свойства кристалла (основа [*] и примесь) | Энергия когезии  | EC*, EC          |
|   | центр d-зоны   | DC*, DC          |
| Свойства поверхности (только основа)      | центры d-зоны верхнего поверхностного слоя и слаба               | DT*, DS*         |
|   | Уровень Ферми слаба  | F*               |

**Рисунок 16 Признаки, используемые для создания более сложных дескрипторов**

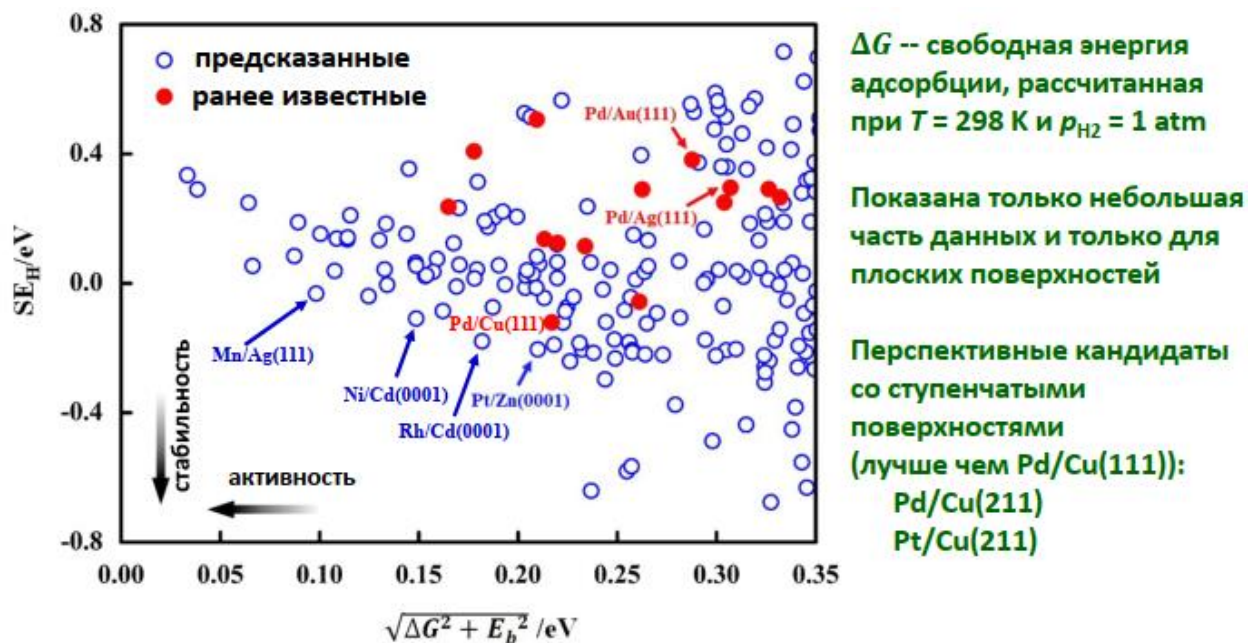


Для расчетов были использованы: полноэлектронный полнопотенциальный DFT-код с FHI-aims; функционал RPBE, выбранный на основе сравнения экспериментально полученных значений энергий связи атомов водорода с поверхностью для чистых металлов  $BE_H$ ; расчеты барьера  $E_b$  были проведены методом упругой ленты с взбирающимися точками (climbing-image nudged elastic band); тренировочные данные включали примерно 300 SAAC различных одноатомных катализаторов: для них были рассчитаны энергии связи, энергии сегрегации (как в присутствии водорода, так и без),  $BE_H$ , SE ( $SE_H$ ), а также энергетические барьеры  $E_b$ ; основные признаки были рассчитаны для  $>5\ 000$  SAAC кандидатов. Полученные результаты дескриптора для разных свойств представлены на рис. 17. Модель очень сложна, однако она не позволяет определить, какие признаки сильнее влияют на целевое свойство.

| свойство | $d^m$   | дескриптор   | коэффициент  |
|----------|---------|--|--------------|
| $BE_H$   | $d_1^5$ | $(EA^*+2F^*-EC)\cdot DT^*\cdot EH^*/(EC^*+F^*)$                | 0.12653E+00  |
|          | $d_2^5$ | $\sqrt[3]{DC}\cdot H^*\cdot DT^*\cdot ( EA^*-EH^* - EC-EC^* )$ | -0.20440E-02 |
|          | $d_3^5$ | $ EH^*-L^*- EH-F^*  /(DC^2+EC\cdot EC^*)$                      | -0.50891E+00 |
|          | $d_4^5$ | $ EH-F^*-EH^* - EC^*-EC- DT^*-F^*  $                           | 0.34705E-01  |
|          | $d_5^5$ | $L\cdot EC\cdot (EA^*+DS^*- H-EH / L^*-EH^* $                  | -0.48772E-04 |
| $E_b$    | $d_1^6$ | $((IP^*-L)- EC^*-DT^* )/ EC/DC-L^*/IP^* $                      | -0.87339E-01 |
|          | $d_2^6$ | $(EA^*+DC^*+ DC-DT^* )/(EA^*+EH^*+ L^*-F^* )$                  | -0.19577E-01 |
|          | $d_3^6$ | $(DC+EH^*)\cdot (EC^*-F^*)\cdot ( L-EC - EC-EH )$              | -0.13173E-01 |
|          | $d_4^6$ | $(DT^*-EH)\cdot DC\cdot (H/EC+EA^*/L^*)/EC^*$                  | -0.19172E-01 |
|          | $d_5^6$ | $e^{EC}\cdot EH\cdot DS^*/((L^*-DS^*)+ H^*-EC^* )$             | 0.33549E-01  |
|          | $d_6^6$ | $DC^2\cdot (EC^*-F^*)/(DT^*-F^*-EA+EC)$                        | -0.14362E-02 |
| SE       | $d_1^6$ | $(EC+IP+ F^*-DT^* )/(IP^*/R+H^*/\underline{dd}^*)$             | -0.82665E+00 |
|          | $d_2^6$ | $ DC-EB^* \cdot (L-DC-EC)/EB^2$                                | 0.30742E+00  |
|          | $d_3^6$ | $  EC^*-L^*+ DC-DS^* - DC-F^* - EC-F^*  $                      | 0.11317E+00  |
|          | $d_4^6$ | $ H-IP-L+IP^* /((DC/EC)+(EC/H))$                               | 0.17455E+00  |
|          | $d_5^6$ | $(F^*-EC)\cdot (L^*-DT^*-IP)/(F^*-EB^*)$                       | -0.51761E-02 |
|          | $d_6^6$ | $EC^*\cdot DC\cdot (EB^*-L)\cdot (L+L^*-EC-DS^*)$              | -0.80032E-03 |

Рисунок 17. Сложные дескрипторы, используемые в расчетах

Результат высокопроизводительного поиска показан на рис. 18.



**Рисунок 18. Результаты высокопроизводительного поиска**

Можно сделать вывод о сложности нахождения катализатора, проявляющего одновременно и стабильность, и высокую активность. Красными точками показаны ранее известные катализаторы, например, палладий, золото. С помощью высокопроизводительного поиска удалось найти несколько других кандидатов. Например, ожидается, что марганец на поверхности серебра будет хорошим катализатором.

Для интерпретации полученных результатов было предложено использовать метод обнаружения подгрупп (рис. 19).



**Какие признаки наиболее важны для низкого барьера, высокой стабильности и удовлетворения принципа Сабатье?**

**Обнаружение подгрупп (SGD) — поиск скрытых локальных структур в данных**

**Модели SISSO → значения целевого свойства, основные признаки**

- 1) Для каждого основного компонента выполнить кластерный анализ k-средних значений
- 2) найти выражения типа (признак1 > a1) AND (признак2 < a2) AND ... такие, что

**а) Ширина распределения целевого свойства в подгруппе невелика**

**б) Количество элементов в подгруппе максимально**

**с) среднее значение целевого свойства для подгруппы является оптимальным (например, минимальный реакционный барьер, минимальная энергия сегрегации)**

Friedman, J.H. & Fisher, N.I., *Statistics and Computing* **9**, 123-143 (1999);

Boley, M et al., *Data Mining and Knowledge Discovery* **31**, 1391-1418 (2017)

**Рисунок 19. Поиск подгрупп**

Для обнаружения подгрупп (SGD) берутся значение целевого свойства и основные признаки из модели SISSO. Затем для каждого основного компонента выполняется кластерный анализ k-средних значений. Далее проводится метод обнаружения подгрупп с учетом следующих замечаний:

а) ширина распределения целевого свойства в подгруппе должна быть небольшой;

б) количество элементов в подгруппе должно быть максимально;

с) среднее значение целевого свойства для подгруппы должно быть оптимальным (например, минимальный реакционный барьер, минимальная энергия сегрегации).

Результаты представлены на рис. 20. Энергия когезии для добавленного элемента должна быть не очень низкой, то есть связи между атомами поверхности должны быть сильнее, чем связи между поверхностью и дополнительным атомом (чтобы поверхность его как бы выталкивала). Также есть дополнительное условие – энергия ионизации диспергированного элемента должна быть высокой, чтобы избежать углубление атома в поверхность.

**Подгруппа, минимизирующая энергию сегрегации SE:**

$$(EC^* \leq -3.85 \text{ eV}) \text{ AND } (EC > -3.36 \text{ eV}) \text{ AND } (IP \geq 7.45 \text{ eV})$$

**Подгруппа, минимизирующая  $\sqrt{\Delta G^2 + E_b^2}$ :**

$$\underbrace{(-2.85 \text{ eV} \leq DC \leq 1.95 \text{ eV})}_{\text{барьер}} \text{ AND } \underbrace{(DT^* \leq -0.17 \text{ eV})}_{\text{связывание}}$$

**Подгруппа с максимальным влиянием адсорбции H на SE (минимальная  $SE_H - SE$ ):**

$$\underbrace{(EH^* \geq -2.35 \text{ eV})}_{\text{связь H с металлом-основанием не слишком сильная}} \text{ AND } \underbrace{(EC^* > -2.73 \text{ eV}) \text{ AND } (EC < -5.98 \text{ eV})}_{\text{условие, обратное по отношению к SE}}$$

связь H с металлом-основанием  
не слишком сильная

условие, обратное по отношению к SE

**Рисунок 20. Поиск подгрупп**

Далее была определена подгруппа, минимизирующая комбинированный показатель для принципа Сабатье и энергетический барьер. Затем были проанализированы подгруппы по  $\Delta G$  и высоте барьера. Было установлено, что в общей подгруппе центр d-зоны диспергированного элемента должен быть не слишком низок и не слишком высок, а центр d-зоны для основной поверхности должен быть не слишком высок. Эти условия отвечают за высоту барьера и силу связывание.

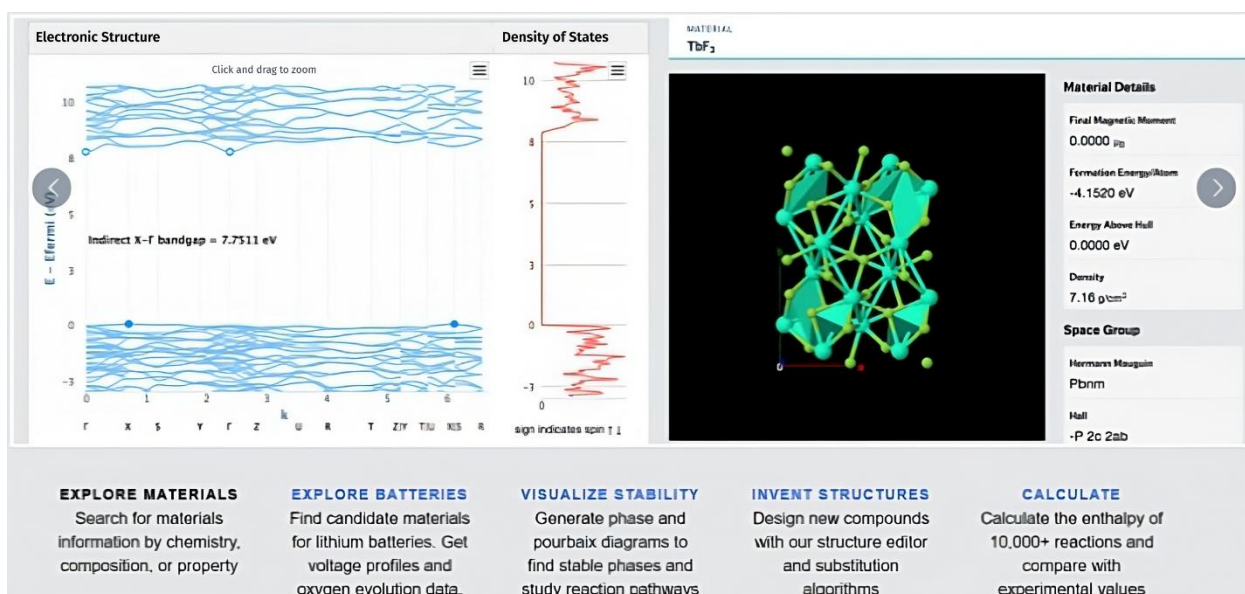
Еще была выделена подгруппа с максимальным влиянием адсорбции атома водорода H на энергию сегрегации SE. Она находилась путем обнаружения подгрупп со свойством разности между энергиями сегрегацией с водородом и без. Видим, что для эффективного взаимодействия атома водорода с диспергированным элементом его связь с металлом-основанием должна быть относительно слабой. Одновременно стоит обратное условие по отношению к энергии сегрегации. Для решения задачи необходимо подобрать оптимальное соотношение этих показателей.

Программное обеспечение SISSO и SGD можно найти по следующим ссылкам:

SISSO: <https://github.com/rouyang2017/SISSO>.

SGD: <https://bitbucket.org/realKD/creedo/wiki/Home>.

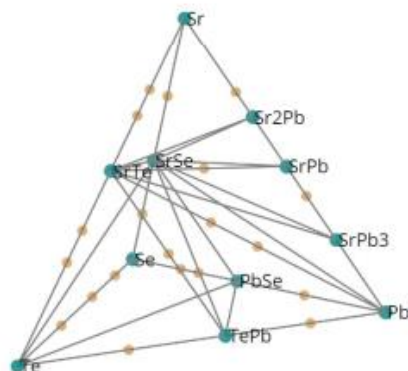
Поговорим о вычислительных базах данных. Их идея состоит в создании инфраструктуры для хранения и анализа полученных данных. Одной из самых известных баз данных материалов является “The Materials Project”. Их девиз: «Мы полностью контролируем данные. Вы можете внести свой вклад, но сначала согласуйте результаты с нами». База данных позволяет (рис. 21) найти кристаллическую структуру, зонную структуру, энергию образования и многое другое. Все расчеты выполнены в рамках обобщенного градиентного приближения теории функционала плотности или DFT + U.



**Рисунок 21. База данных “The Materials Project”**

Также существует база данных OQMD (The Open Quantum Materials Database) (рис. 22). Она содержит не только экспериментально известные структуры, но и гипотетически возможные, которые не были получены экспериментально. Все расчеты выполнены с помощью GGA или GGA+U. Позволяет найти энергии образования и декомпозиции и др.

The OQMD is a database of DFT calculated thermodynamic and structural properties of **815,654** materials, created in Chris Wolverton's group at Northwestern University.



### Shortcuts

#### Search

Material Compositions

#### Query

Materials Data

#### Create

Phase Diagrams

#### Determine

Ground State  
Compositions (GCLP)

#### Visualize

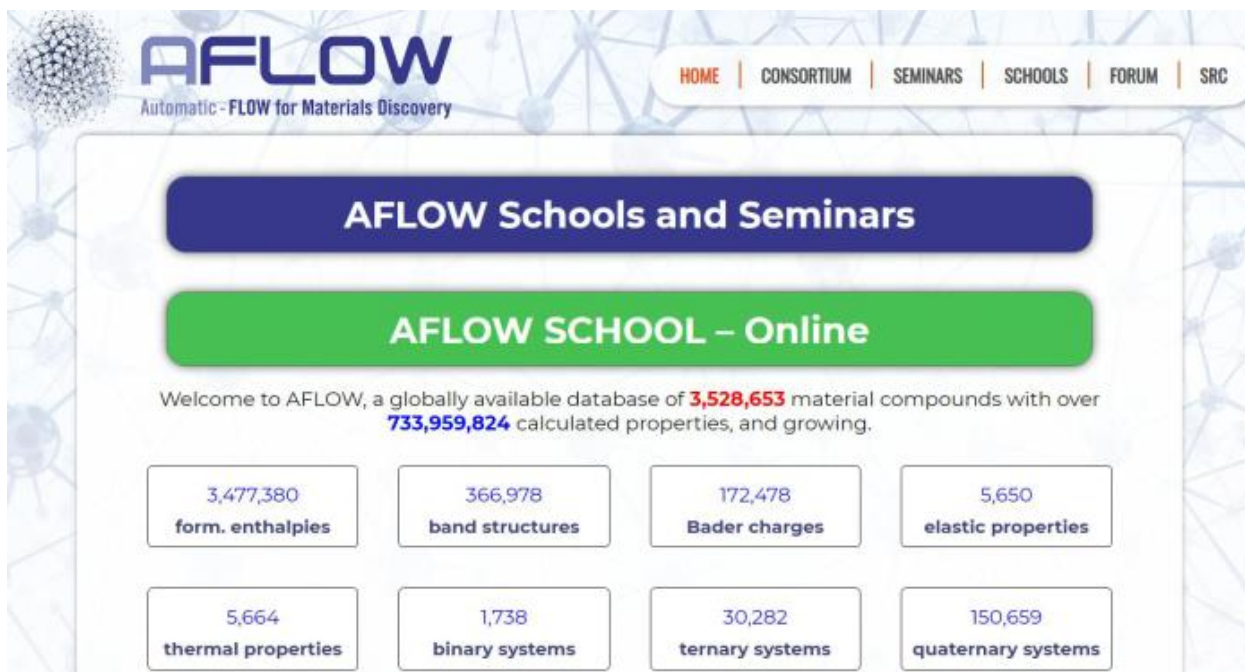
Crystal Structures

#### RESTful API

OQMD API  
Optimade API

**Рисунок 22. База данных OQMD**

Активно развивающейся базой данных является AFLOW (рис. 23). Расчеты выполнены с помощью GGA, GGA+U, ACBN0 (псевдогибридный функционал, который по скорости не уступает другим методам и является более точным за счет автоматического вычисления поправок для разных атомов). Позволяет найти электронные и фоновые спектры, магнитные и термодинамические свойства материалов, и др.



*Рисунок 23. База данных AFLOW*

Одной из самых известных баз данных в Европе является NOMAD (Novel Materials Discovery) Centre of Excellence (рис. 24). Основное отличие от других баз заключается в хранении данных пользователей. Девиз: Мы будем хранить все ваши данные; даже неточные вычисления имеют ценность (например, они могут быть использованы для обучения). Условие – должны быть известны все метаданные (версия кода, метод, базис и т. д.). Включает данные из AFLOW, OQMD, Materials Project. Позволяет найти оптимизированные геометрии, термодинамические свойства материалов и др.



# NOMAD Centre of Excellence

Bringing computational materials science to exascale

## Exascale Codes

- Bringing DFT, Green-function methods, and coupled-cluster theory to exascale
- Supporting entire code families, covering planewaves (PW), linearized augmented PWs, and atom-centred orbitals
- Follow us on [GitHub](#)

## Exascale Workflows

- Enabling exascale computations by advanced workflows
- Covering high-throughput computations and beyond-DFT workflows
- Learn how to work with ASE/ASR and FireWorks in this [tutorial](#)

## Extreme-scale data





- Advance the [NOMAD AI toolkit](#) and bring it towards near-real-time performance
- Like to visit the [NOMAD Laboratory](#) and its services for up- and downloading, and exploring materials data?
- Watch our [video tutorials](#) to learn how to work with the AI toolkit

***Рисунок 24. База данных NOMAD Centre of Excellence***

Также интересен проект AiiDA – это набор средств для создания и управления собственной базой данных. Девиз данного проекта: Создайте свою собственную базу данных с помощью наших инструментов. Также предоставляет средства для выполнения высокопроизводительных вычислений (рис. 25).

# AiiDA

Automated Interactive Infrastructure and Database for Computational Science

-  Workflows
-  Data provenance
-  Plugin framework
-  HPC Interface
-  Open Science
-  Open source

If you use AiiDA please cite:

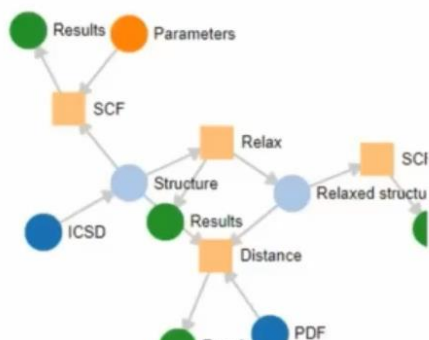
AiiDA 1.x

Main paper: S.P. Huber et al., *Scientific Data* 7, 300 (2020)

AiiDA engine: M. Uhrin et al., *Comp. Mat. Sci.* 187 (2021)

AiiDA 0.x

First paper, ADES model: G. Pizzi et al. *Comp. Mat. Sci.* 111, 218-230 (2016) (open access version)



## What is AiiDA?

AiiDA is an open-source Python infrastructure to help researchers with automating, managing, persisting, sharing and reproducing the complex workflows associated with modern computational science and all associated data.

AiiDA is built to support and streamline the four core pillars of the **ADES model**: Automation, Data, Environment, and Sharing. Key features include:

- Workflows:** AiiDA allows to build and execute complex, auto-documenting workflows linked to multiple codes on local and remote computers.
- High-throughput:** AiiDA's event-based workflow engine supports tens of thousands of processes per node with full check-pointing.
- Data provenance:** AiiDA automatically tracks and records inputs, outputs and metadata of all calculations and workflows in extensive provenance graphs that preserve the full lineage of all data.
- Advanced queries:** AiiDA's query language enables fast graph queries on millions of nodes.
- Plugin interface:** AiiDA can support via plugins any computational code and data analytics tool, data type, scheduler, connection mode, etc. (see [public plugin repository](#))
- HPC interface:** AiiDA can seamlessly deal with heterogeneous and remote computing resources; it works with many schedulers out of the box (SLURM, PBS Pro, torque, SGE or LSF).
- Open science:** AiiDA allows to export both full databases and selected subsets, to be shared with collaborators or made available and browsable online on the [Archive](#) and [Explore](#) sections of [Materials Cloud](#).
- Open source:** AiiDA is released under the MIT open-source license.

## Most recent news

AiiDA features in Nature Reviews Physics "Tools of the Trade"

To mark the release of AiiDA v2.0, it was presented as a "Tool of the..."

[Read More](#)

AiiDA v2.0.0 released

Dear AiiDA users, We are very happy to announce that today we released the final...

[Read More](#)

*Рисунок 25. Проект AiiDA*