

Возможность определения локальной атомной структуры цеолита Cu-MOR с использованием XANES-спектроскопии и методов машинного обучения

© Я.Н. Гладченко-Джевелекис, Г.Б. Сухарина, А.М. Ермакова, К.Д. Кулаев, В.В. Прядченко, Е.Е. Поносова, Э.И. Шеметова, Л.А. Авакян, Л.А. Бугаев

Южный федеральный университет,
344090 Ростов-на-Дону, Россия
e-mail: ygl@sfedu.ru, gbsukharina@sfedu.ru

Поступило в Редакцию 16 ноября 2023 г.
В окончательной редакции 31 декабря 2023 г.
Принято к публикации 26 января 2024 г.

Проанализированы способы определения локальной структуры центров меди цеолитов с использованием комбинации методов машинного обучения и спектроскопии рентгеновского поглощения. Сложность определения структуры центров меди обусловлена их неравномерным распределением в каркасе цеолита, что затрудняет интерпретацию полученных Cu K-XANES экспериментальных спектров. Алгоритмы машинного обучения с использованием синтетических данных, полученных в программном комплексе FDMNES, позволили установить, что локация центров меди в конкретном кольце цеолита может быть определена с точностью 0.97 согласно метрике F1.

Ключевые слова: цеолиты, атомная структура, XANES, ML-классификация, нейросети.

DOI: 10.61011/JTF.2024.04.57533.287-23

Введение

Цеолиты — это микропористые алюмосиликатные соединения, используемые в промышленности как катализаторы. Например, в С1-химии цеолиты активно исследуют и оптимизируют для использования в процессе конверсии метана в метanol. Высокая катализическая активность цеолитов обусловлена наличием активных центров, содержащих атомы переходных металлов [1–5]. При этом детальное изучение структуры каталитически активных центров представляет собой сложную задачу, поскольку распределение таких центров в каркасе цеолита является неравномерным, так как зависит от множества взаимосвязанных факторов (например, молярного соотношения реагентов при синтезе, времени реакции, ее температуры и т.д.) [6–9]. Таким образом, получение эффективных катализаторов связано с исследованием конкретных структурных параметров металлических центров и их влияния на каталитические свойства, что на сегодняшний день не изучено системно [10] и требует совершенствования существующих подходов решения таких задач.

Методы машинного обучения (МО) являются перспективным инструментом в определении структуры различных материалов [11–13], в частности, цеолитов, за счет возможностей, связанных с расчетами для нелинейных или массивных комбинаторных процессов, которые не решаемы с помощью традиционных подходов [14]. Учитывая множество факторов, влияющих на структуру при синтезе цеолитов, а также многообразие

сформированных типов центров меди, влияющих на каталитические свойства, использование МО особенно актуально для улучшения процесса синтеза. Применение МО в данной области позволит автоматизировать и упростить процесс определения структурных характеристик цеолитов по экспериментальным данным для установления закономерностей влияния структурных особенностей цеолитов на свойства катализатора [15]. Хорошо описанным примером использования МО в данной области является работа, описывающая прогнозирование упругих свойств чистых кремнеземных цеолитов [16]. Авторы использовали Gradient Boosting Regressor (GBR) [17,18] и обучающую выборку упругих свойств, рассчитанную с помощью DFT (Density Functional Theory). Оказалось, что данный подход МО способен прогнозировать модули объемного сжатия и сдвига цеолитов с точностью выше, чем при использовании классических межатомных потенциалов, и за гораздо меньшее время. МО также применялось для прогнозирования возможных промежуточных продуктов реакции преобразования углекислого газа в метанол с использованием цеолитов, модифицированных металлами [19], с помощью алгоритмов XGBoost и ExtraTrees.

Настоящая работа посвящена разработке методики определения локальной атомной структуры центров меди, в частности, кольца в каркасе медьсодержащего цеолита типа морденит (Cu-MOR), в котором расположены атомы меди, включающей методы МО и рентгеновскую спектроскопию XANES (X-ray absorption near edge structure).

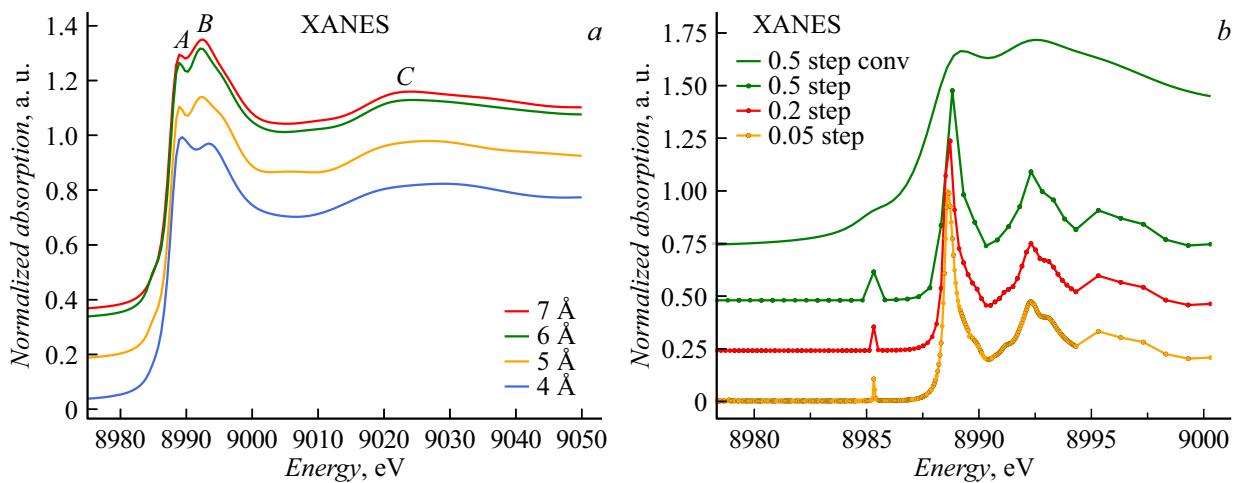


Рис. 1. Теоретические Cu K-XANES-спектры ближнего окружения меди в каркасе цеолита, рассчитанные для кластеров различного размера (a); с различным шагом сетки вблизи края поглощения для кластера 6 Å (b).

1. Методы и подходы

Поход, сочетающий методы МО и спектроскопию рентгеновского поглощения XANES [20,21] с применением синтетических данных позволяет не только рассмотреть большое число моделей центров меди, но и значительно ускорить процесс определения структурных параметров.

Одной из задач машинного обучения является классификация, которая используется для разделения данных на классы и позволяет предсказывать результат на основе входных данных. В настоящей работе классами являются кольца каркаса цеолита, в которых расположены атомы меди, а входными данными, на основе которых делаются предсказания, — Cu K-XANES-спектры различных структур Cu-MOR. Процедура включает обучение с учителем, в процессе которого структурные модели с заранее известными кольцами, в которых находятся центры меди, и значениями Cu K-XANES-спектров используются для обучения классификатора. После этого классификатор можно применять для определения колец, включающих центры меди, по экспериментальным Cu K-XANES-спектрам.

Моделирование структур проводилось с использованием библиотеки ASE (Atomic Simulation Environment) [22]. ASE — это пакет программного обеспечения, написанный на языке программирования Python и предназначенный для настройки, управления и анализа атомистического моделирования. ASE предоставляет модули для выполнения многих стандартных задач моделирования, таких, как оптимизация структуры и молекулярная динамика. В настоящей работе модели центров меди — объекты типа Atoms, представляющие собой совокупность атомов с заданными декартовыми позициями. Геометрическая оптимизация была выполнена методом LAMMPS из библиотеки ASE с использованием потенциала, описанного в работе [23].

Спектры Cu K-XANES для различных структурных моделей центров меди в Cu-MOR были рассчитаны методом конечных разностей, реализованном в программе FDMNES [24]. Метод конечных разностей представляет собой численный метод решения дифференциальных уравнений, в котором решение ищется на сетке точек. В данной задаче решается уравнение Шредингера для сферической области вокруг поглощающего атома, включающей кластер, содержащий необходимое число атомов. Для воспроизведения всех основных особенностей Cu K-XANES-спектров моделей центров меди в Cu-MOR достаточно проводить расчеты для кластера радиусом 6 Å с шагом сетки вблизи края поглощения 0.5 Å. Выбор кластера с указанными параметрами обусловлен результатами проверки влияния соответствующих параметров на форму теоретического Cu K-XANES-спектра.

С учетом размера пор в каркасе цеолита были выполнены расчеты для кластеров ближнего окружения меди различных размеров, а именно 4, 5, 6 и 7 Å. В результате установлено, что насыщение теоретического Cu K-XANES-спектра, необходимое для реализации задачи классификации центра меди в кольце цеолита, наблюдается при использовании кластера размером 6 Å — воспроизведены все основные особенности спектра (A, B, C) для ближнего окружения меди в каркасе Cu-MOR (рис. 1, a). Кроме того, были выполнены расчеты Cu K-XANES-спектров с использованием различного шага сетки 0.5, 0.2 и 0.05 Å вблизи края поглощения, что также не привело к существенным изменениям формы теоретического спектра (рис. 1, b). Расчеты были выполнены в возбужденном состоянии с использованием потенциала Хедина-Ландквиста.

Для возможности графического отображения выборки был использован метод UMAP (Uniform Manifold Approximation and Projection), который является одним из наиболее продвинутых методов нелинейного понижения размерности и визуализации данных. Метод

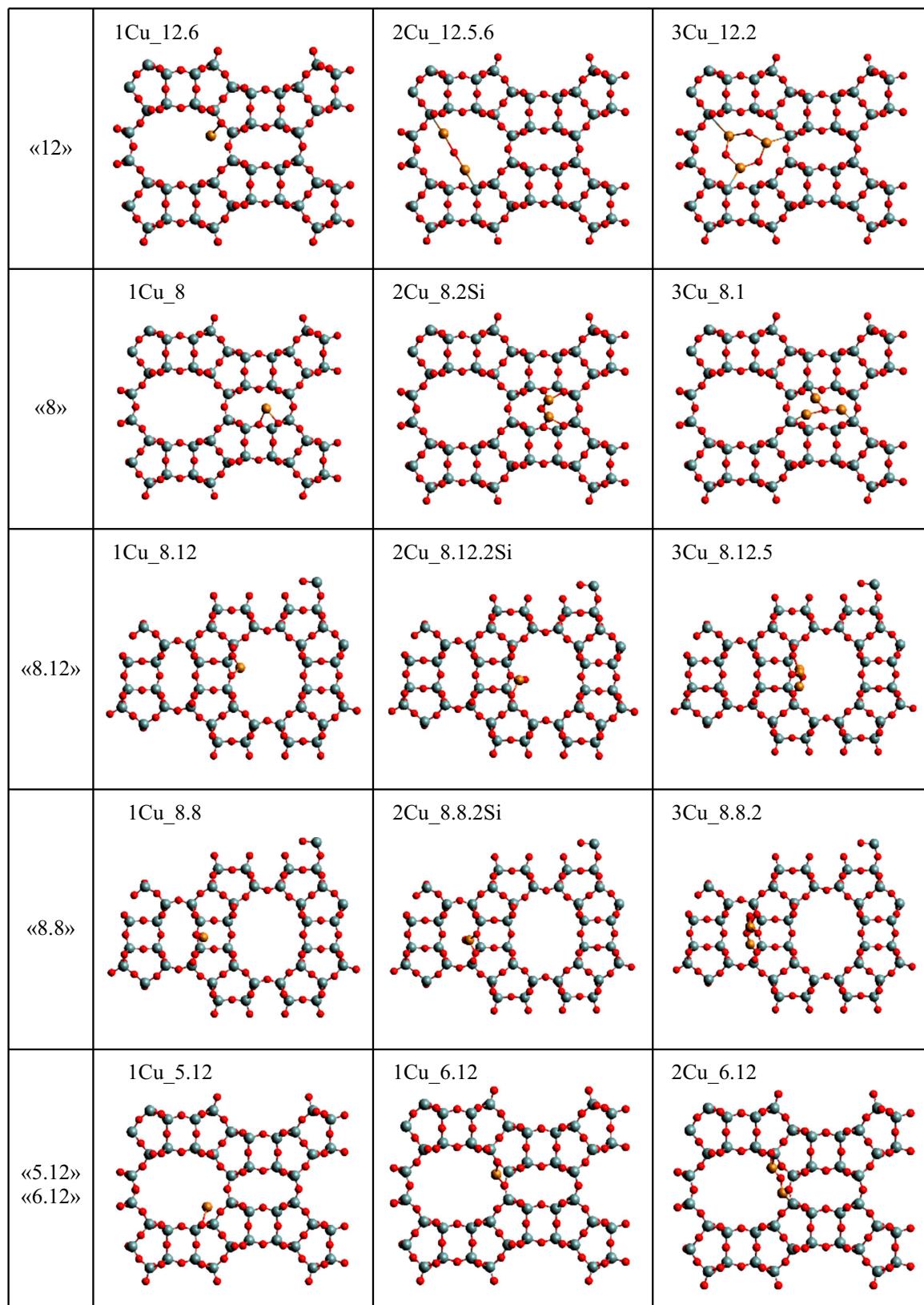


Рис. 2. Схематичное представление некоторых моделей медных центров в каркасе Cu-MOR. Число атомов меди указано перед названием основного класса, например 1Cu_12.6.

Таблица 1. Названия моделей, входящих в выборку, используемую для обучения нейронной сети

Класс	Список объектов, входящих в этот класс		
12	12.5_1Cu, 12.6_1Cu, 12.8_1Cu	12.5.5_2Cu, 12.5.6_2Cu, 12.5.8_2Cu	12.1_3Cu, 12.2_3Cu, 12.3_3Cu
8	8_1Cu	8.2Si_2Cu, 8.3Si_2Cu	8.1_3Cu, 8.2_3Cu
8.12	8.12_1Cu	8.12.2Si_2Cu, 8.12.3Si_2Cu	8.12.1_3Cu, 8.12.2_3Cu, 8.12.3_3Cu, 8.12.4_3Cu, 8.12.5_3Cu, 8.12.6_3Cu
8.8	8.8_1Cu	8.8.2Si_2Cu, 8.8.3Si_2Cu	8.8.1_3Cu, 8.8.2_3Cu 8.8.3_3Cu, 8.8.4_3Cu, 8.8.5_3Cu, 8.8.6_3Cu
6.12	6.12_1Cu	6.12_2Cu	
5.12	5.12_1Cu		

УMAP основан на идеи построения многообразия низкой размерности, сохраняющего топологическую структуру данных высокой размерности, т.е. позволяющий сохранять как локальный характер распределения, так и глобальную структуру исходных данных [25].

2. Результаты и обсуждение

Для применения методов МО к решению задачи определения локальной атомной структуры Cu-MOR была сформирована обучающая выборка, содержащая Cu K-XANES-спектры, рассчитанные в программном комплексе FDMNES для набора моделей центров меди, расположенных в различных кольцах внутри каналов каркаса морденита. Цеолиты имеют сложную алюмосиликатную структуру, поэтому для удобства интерпретации полученных расчетных данных была введена собственная номенклатура колец каркаса Си-морденита (рис. 2).

Перечень моделей был сформирован на основе имеющихся на сегодняшний день литературных данных [26–28], в которых было учтено наиболее вероятное расположение меди в рассматриваемых кольцах цеолита. На рис. 2 и 3 схематично представлены примеры моделей центров меди, расположенных в различных кольцах цеолита и содержащих 1, 2 или 3 атома меди. Названия основных классов для моделей были выбраны исходя из количества атомов кремния в кольце или канале, в котором располагаются атомы меди: „12“ — атомы меди находятся в 12-ричном канале; „8“ — атомы меди находятся в 8-ричном канале; „8.12“ и „8.8“ — атомы меди находятся в 8-ричном кольце, которое является элементом 12-ричного или 8-ричного канала соответственно; „5.12“ и „6.12“ — атомы меди находятся в 5-ричном или 6-ричном кольце 12-ричного канала.

Далее основные классы разбивались на подклассы, в которых положения меди различались друг от друга в

рамках главного класса. Для класса „12“ при рассмотрении одного атома меди были выбраны три подкласса — „12.5“, „12.6“, „12.8“. Второй индекс ставится в зависимости от близости нахождения к тому или иному кольцу в 12-ричном канале. При рассмотрении этого же класса, но с двумя атомами меди, были рассмотрены также три подкласса — „12.5.5“, „12.5.6“, „12.5.8“. На рис. 2 показан пример для модели „12.5.6“. В ней атомы меди соединены друг с другом через атом кислорода, а с каркасом — через атомы алюминия, входящие в 5-ричное и 6-ричное кольцо в 12-ричном канале. Таким же образом для двух атомов меди в кольцах „8“, „8.12“ и „8.8“ модели разбиваются на два подкласса: „8.2Si“, „8.3Si“; „8.12.2Si“, „8.12.3Si“; „8.8.2Si“, „8.8.3Si“, где указано количество атомов кремния в кольце между атомами алюминия (рис. 3).

Для трех атомов меди в канале „12“ рассматривались три вида моделей: „12.1“, „12.2“, „12.3“, которые отличались положениями атомов меди в кольце и связями через кислород между собой и с каркасом цеолита. Классификация названий для всех остальных моделей вводилась аналогичным образом: для трех атомов меди в канале „8“ рассматривались модели „8.1“ и „8.2“; в кольце „8.12“ — модели „8.12.1“, „8.12.2“, „8.12.3“, „8.12.4“, „8.12.5“, „8.12.6“; в кольце „8.8“ — модели „8.8.1“, „8.8.2“, „8.8.3“, „8.8.4“, „8.8.5“, „8.8.6“.

Таким образом, при рассмотрении каналов „12“ и „8“ (например, модель „12.5“), первая цифра — это канал, а вторая — кольцо, такие обозначения являются целесообразными, поскольку положение атома меди рассматриваемого медного центра варьируется вдоль канала. При рассмотрении остальных классов используется другая сквозная нумерация, в которой первая цифра — это кольцо, а вторая — канал, к которому это кольцо принадлежит. В данном случае значимым элементом является кольцо, которое представляет собой искомый класс. В табл. 1 представлены названия всех классов, входящих в выборку.

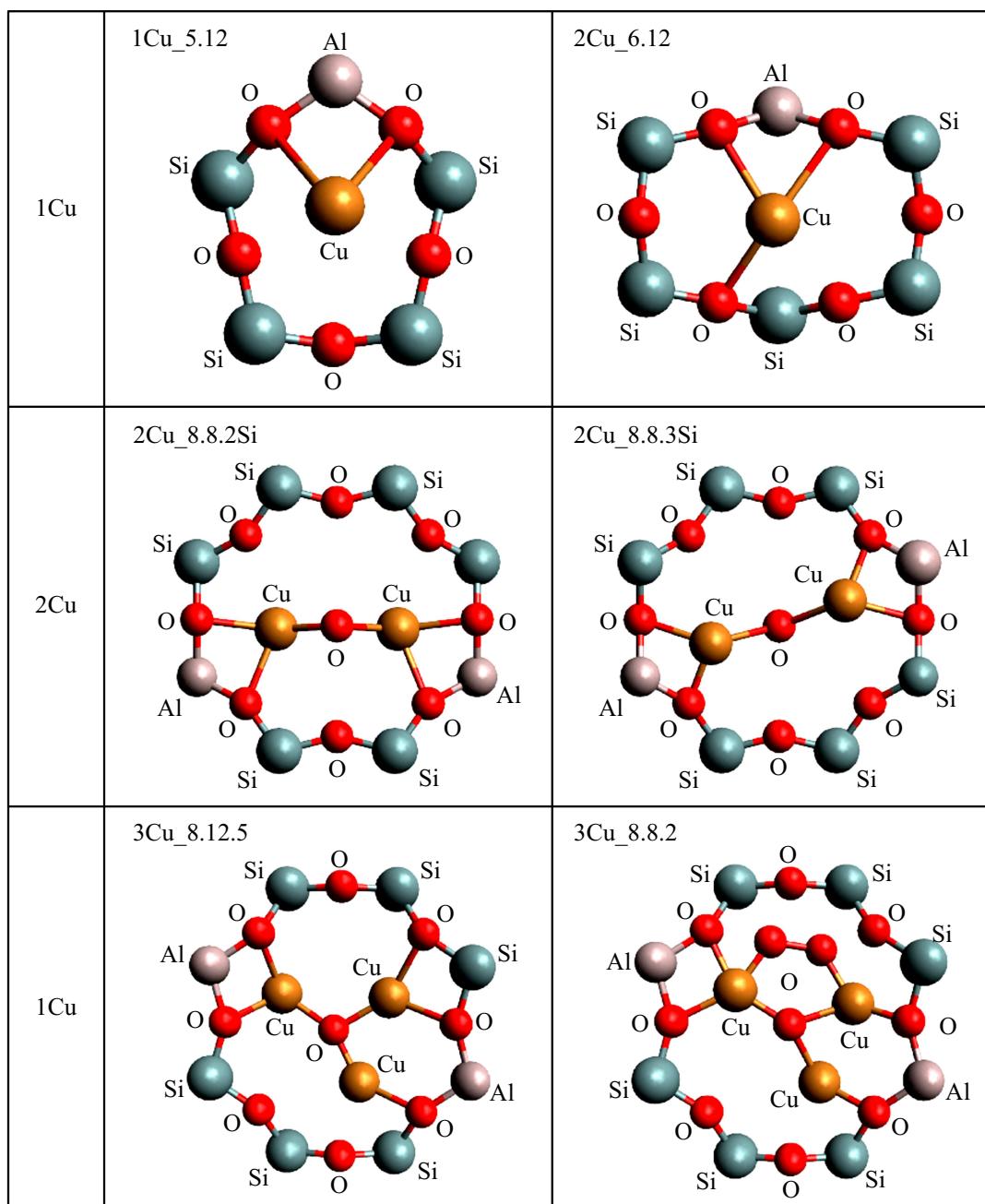


Рис. 3. Схематичное представление некоторых колец в каркасе Cu-MOR, в которых возможно формирование центров меди, содержащих 1, 2 и 3 атома меди соответственно.

Для обучения и тестирования классификатора был сформирован набор данных, включающий Cu K-XANES-спектры для каждой модели из имеющейся выборки с различными изменяющимися структурными параметрами ближнего окружения меди: межатомными расстояниями — $R_{\text{Cu-Al}}$ с учетом различных направлений сдвига атома меди и углом поворота φ (рис. 4, c). Перечень интервалов изменения структурных параметров приведен в табл. 2. В результате сформированный набор данных включал 2100 рассчитанных Cu K-XANES-спектров для различных моделей центров меди. При

помещении итогового спектра в выборку были учтены все неэквивалентные положения атомов меди.

Кроме того, была выполнена аугментация теоретических данных с учетом присутствия влияния оксидов меди (CuO и Cu_2O). Для этого в теоретический набор данных были добавлены Cu K-XANES-спектры, представляющие собой линейную комбинацию рассматриваемых моделей центров меди и оксидов меди CuO и Cu_2O с различными вкладами: 25, 50 и 75% для каждого оксида отдельно. Таким образом, выборка составила 14700 теоретических Cu K-XANES-спектров. На рис. 4 в

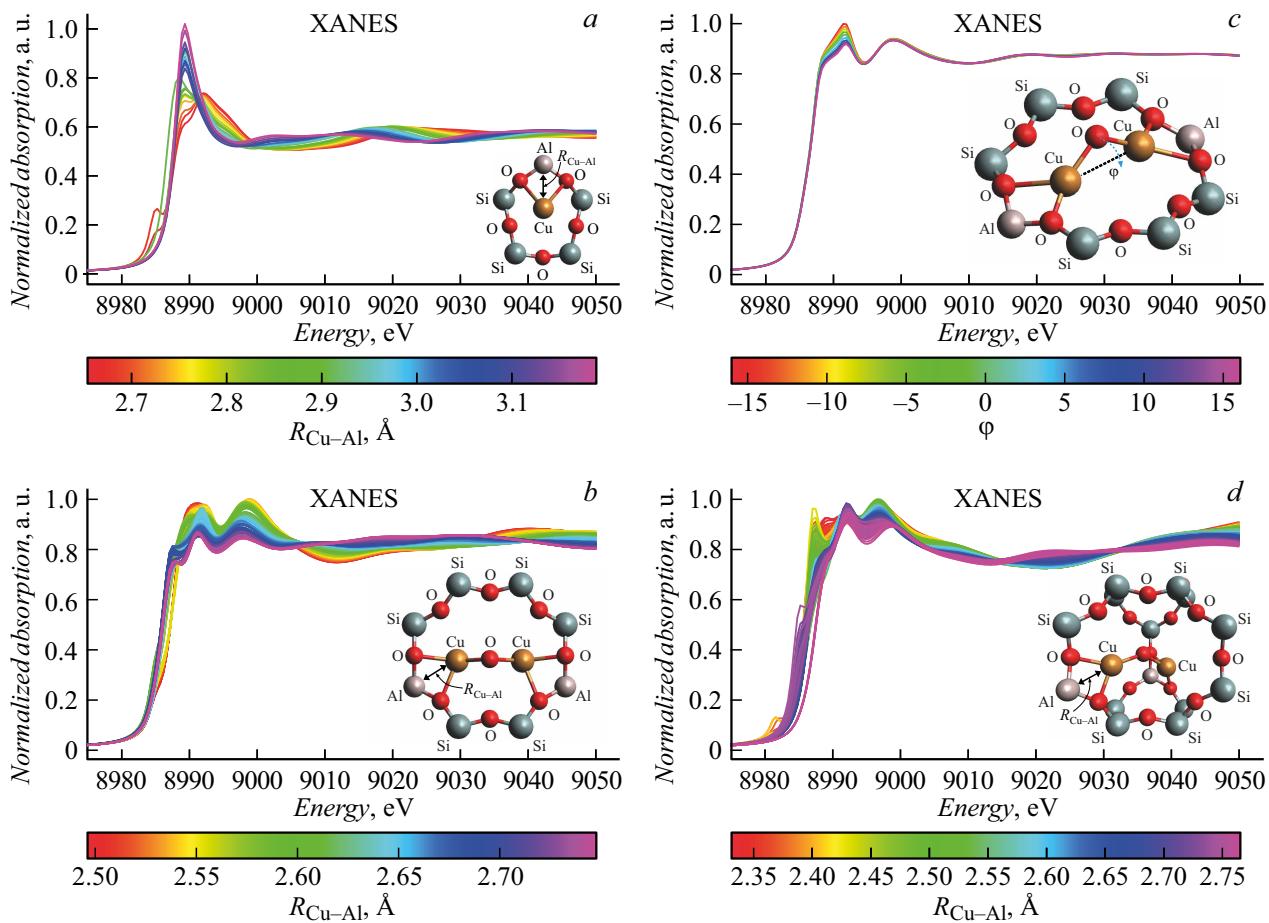


Рис. 4. Рассчитанные Cu K-XANES-спектры для: модели медного центра в кольце 5.12, содержащем 5 атомов кислорода, 4 атома кремния и 1 атом алюминия (a); модели медного центра в кольце 8.8.2Si, содержащем 8 атомов кислорода, 6 атомов кремния и 2 атома алюминия (b); модели медного центра 8.12.3Si в кольце 8.12, где варьируемым параметром является угол поворота φ цепочки Cu-O-Cu на 16° с шагом 4° (c); модели медного центра 8.3Si в кольце 8 (d).

качестве примера представлены элементы выборки для колец, обозначенных согласно введенной номенклатуре „5.12“ (рис. 4, a), 8.8.2Si (рис. 4, b), 8.12.3Si (рис. 4, c) и 8.3Si (рис. 4, d).

В качестве классификатора была выбрана нейронная сеть (MLP — многослойный перцептрон), поскольку она показала наиболее высокую точность, в отличие от классических методов машинного обучения, таких, как Extra Trees Classifier, KNeighbors, Random Forest, для которых метрика F1 составила 0.80, 0.65 и 0.79 соответственно. Перед обучением нейронной сети был проведен анализ данных на возможность их кластеризации с помощью метода нелинейного понижения размерности UMAP. Благодаря понижению размерности до двух параметров стало возможно графическое отображение выборки (рис. 5). При количестве соседей, равном 100, и минимальном расстоянии, равном 1, видно, что данные разделяются на кластеры.

Для определения кольца, содержащего центр меди, использовалась полносвязная нейронная сеть, созданная на основе модулей библиотеки pytorch [29]. Нейронная

сеть состояла из входного линейного слоя, dropout-слоя, слоя ReLU активации, скрытого линейного слоя и softmax-слоя. Метод dropout обнуляет некоторые элементы входного тензора с вероятностью 0.25, используя выборки из распределения Бернулли. Функция softmax изменяет выходной тензор так, чтобы его элементы лежали в интервале $[0, 1]$ и в сумме давали единицу. Значения на выходе после функции softmax можно считать приближенными вероятностями того, что активный медный центр находится в указанных кольцах. В качестве функции потерь использовалась многоклассовая кросс-энтропия. Тестовая выборка составила 33% от общего набора данных. Для реализации процесса тренировки был использован оптимизатор Adam с коэффициентом скорости обучения (learning rate) — 0.01. Спустя 1000 эпох (циклов обучения) точность метрики F1 перестает расти и останавливается на значении 96.6% для тестовой выборки и 97.4% для тренировочной (рис. 6, 7). Важно подчеркнуть, что на первом этапе работы была получена модель, которая обучалась на неаугментированных данных, что необходимо для исключения попадания в тре-

Таблица 2. Вариация структурных параметров при формировании базы данных, $R_{\text{Cu-Al}}$ — расстояния между атомом меди и алюминия, φ — углы поворота

Cu	Модель	Структурный параметр	Изменение параметров, $R_{\text{Cu-Al}}$, ангстремы, и φ , градусы
1	12	$R_{\text{Cu-Al}}$	2.12–3.32
	12.6	$R_{\text{Cu-Al}}$	2.35–3.37
		$R_{\text{Cu-Al}}$	1.99–3.41
		$R_{\text{Cu-Al}}$	3.17–3.89
		$R_{\text{Cu-Al}}$	2.40–4.05
		$R_{\text{Cu-Al}}$	2.36–4.03
	6.12	$R_{1\text{-Cu-Al}}$	2.51–3.15
		$R_{2\text{-Cu-Al}}$	2.53–3.42
	5.12	$R_{\text{Cu-Al}1}$	2.64–3.19
		$R_{\text{Cu-Al}2}$	2.91–3.59
		$R_{\text{Cu-Al}3}$	3.31–3.70
2	12.5.5	$R_{\text{Cu-Al}}$	2.12–3.32
	12.5.6	$R_{\text{Cu-Al}}$	1.92–3.12
	12.5.8	$R_{\text{Cu-Al}}$	2.29–3.49
	6.12	$R_{\text{Cu-Al}}$	1.82–2.82
	8.12.2Si	$R_{\text{Cu-Al}}$	2.48–2.66
		φ	–16 – +16
	8.12.3Si	$R_{\text{Cu-Al}}$	2.51–2.67
		φ	–16 – +16
	8.2Si	$R_{\text{Cu-Al}}$	2.33–2.76
		φ	–16 – +16
3	8.3Si	$R_{\text{Cu-Al}}$	2.33–2.73
		φ	–16 – +16
	8.8.2Si	$R_{\text{Cu-Al}}$	2.5–2.75
		φ	–16 – +16
8.8.3Si	8.8.3Si	$R_{\text{Cu-Al}}$	2.53–2.76
		φ	–16 – +16

нировочную и тестовую выборку похожих спектров. При этом точность модели была оценена при помощи кросс-валидации на 10 подгруппах и составила 89% согласно средней метрике F1. Таким образом, модель является статистически устойчивой и точность ее работы сопоставима с моделью, обученной на аугментированных данных.

Исходя из матрицы ошибок (рис. 7), можно сделать вывод, что нейронная сеть с хорошей точностью, согласно метрике F1, отличает различные структурные кольца цеолита друг от друга. Диагональные элементы представляют количество правильно классифицированных

Таблица 2 (продолжение).

Cu	Модель	Структурный параметр	Изменение параметров, $R_{\text{Cu-Al}}$, ангстремы, и φ , градусы
3	12.1	$R_{\text{Cu-Al}}$	3.73–4.04
	12.2	$R_{\text{Cu-Al}}$	3.73–4.04
	12.3	$R_{\text{Cu-Al}}$	3.34–3.69
	8.1	$R_{\text{Cu-Al}}$	3.37–3.91
	8.12.1	$R_{\text{Cu-Al}}$	2.03–2.55
	8.12.2	$R_{\text{Cu-Al}}$	2.25–2.73
	8.12.3	$R_{\text{Cu-Al}}$	2.03–2.55
	8.12.4	$R_{\text{Cu-Al}}$	2.03–2.55
	8.12.5	$R_{\text{Cu-Al}}$	2.03–2.55
	8.12.6	$R_{\text{Cu-Al}}$	2.03–2.55
8.2	8.2	$R_{\text{Cu-Al}}$	3.62–4.12
	8.8.1	$R_{\text{Cu-Al}}$	1.81–1.92
	8.8.2	$R_{\text{Cu-Al}}$	2.36–2.53
	8.8.3	$R_{\text{Cu-Al}}$	1.81–1.92
	8.8.4	$R_{\text{Cu-Al}}$	1.81–1.92
	8.8.5	$R_{\text{Cu-Al}}$	1.81–1.92
8.8.6	8.8.6	$R_{\text{Cu-Al}}$	2.47–2.56

кольц для каждого класса, а недиагональные элементы представляют неправильно классифицированные кольца.

Заключение

В работе для определения локальной атомной структуры центров меди, сформированных в каркасе цеолита Cu-MOR, мы предложили подход, сочетающий спектроскопию рентгеновского поглощения XANES, компьютерное моделирование и МО. Для полученного набора моделей центров меди различного типа были выполнены расчеты Cu K-XANES-спектров с использованием программного комплекса FDMNES. В результате была предложена собственная подробная номенклатура моделей центров меди, расположенных в различных кольцах цеолита типа морденит и сформирован обучающий набор данных, содержащий 14700 теоретических Cu K-XANES-спектров, с учетом влияния оксидов CuO и Cu₂O. Применение методов кластеризации к полученным модельным данным показало, что Cu K-XANES-спектры, соответствующие моделям центров меди, находящихся в одном структурном фрагменте, локализуются в одной области пространства признаков.

Возможность разбиения признаков по кластерам позволяет определить на основании Cu K-XANES-спектров

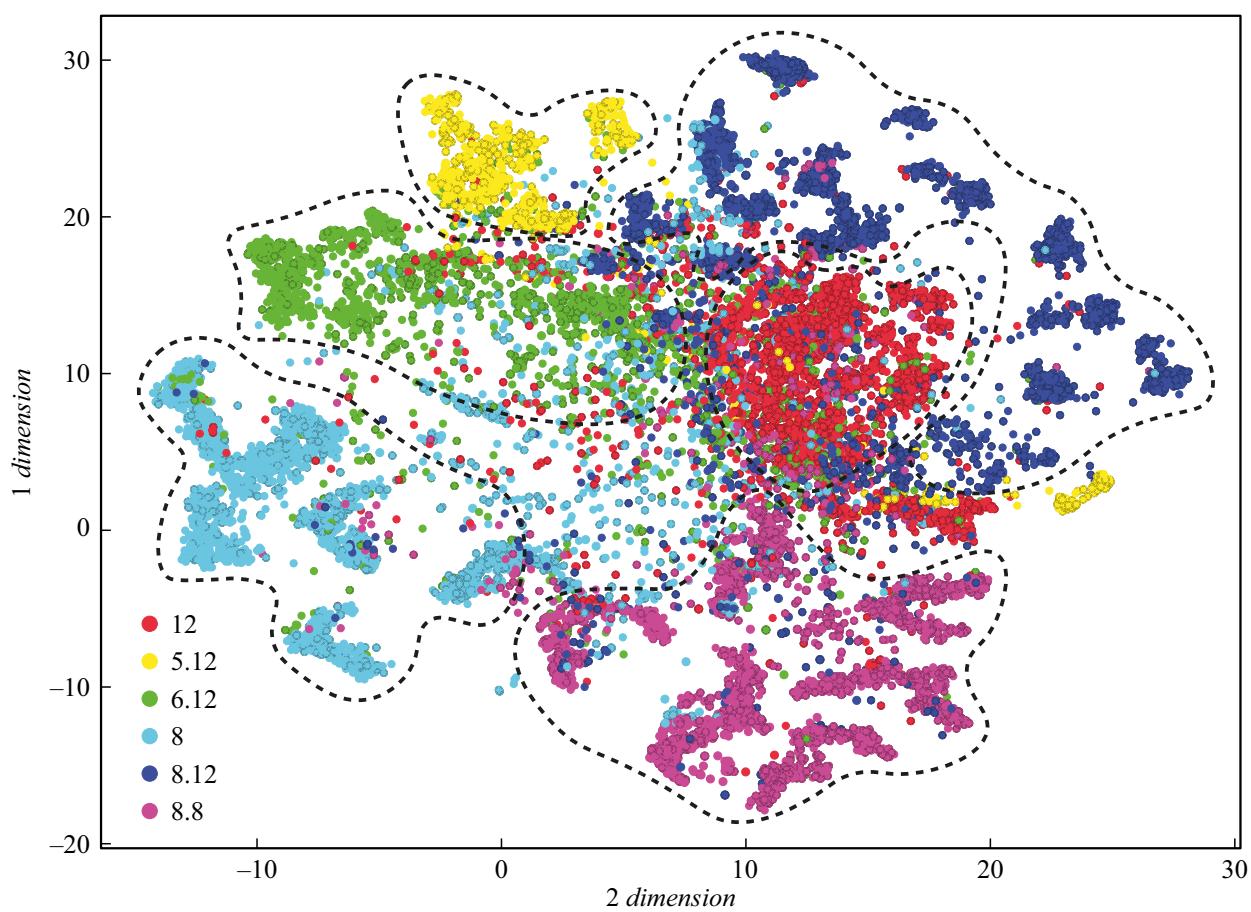


Рис. 5. УМП-отображение, полученное с использованием пакета Python umap для набора данных Cu K-XANES, рассчитанных для моделей медных центров в Cu-MOR. Цвета представляют различные кольца, в которых находится медный центр, согласно введенной номенклатуре (рис. 1).

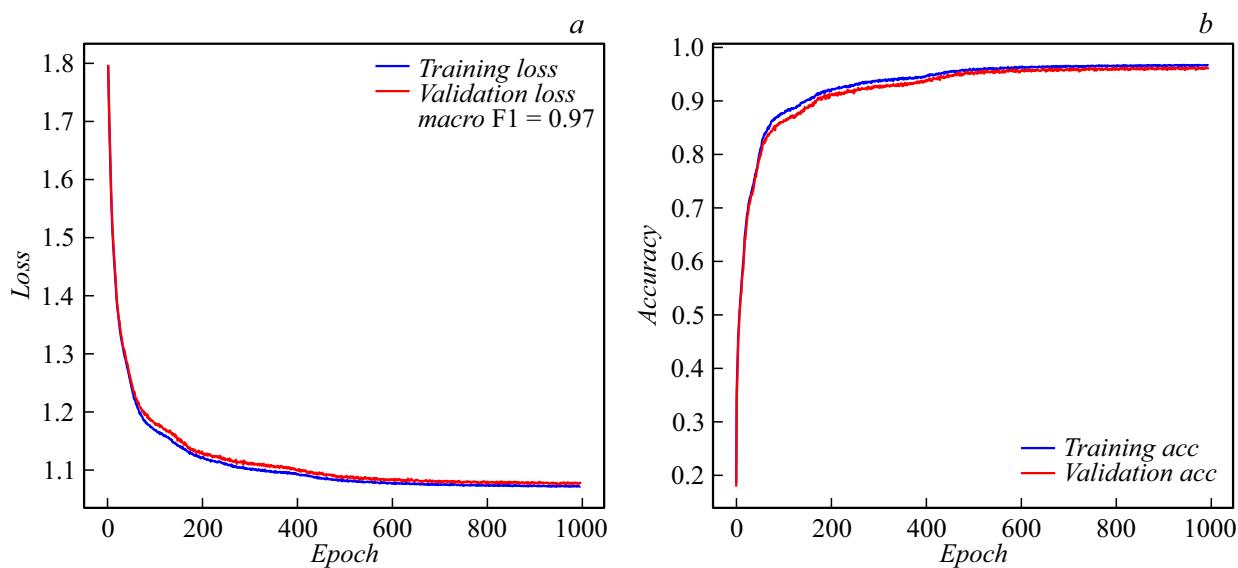


Рис. 6. График функции потерь для нейронной сети, где ось ОХ показывает количество эпох, а ось ОY показывает потери кросс-энтропии (a). График метрики accuracy для нейронной сети в процессе обучения на наборе данных Cu K-XANES-спектров для Cu-MOR (b).

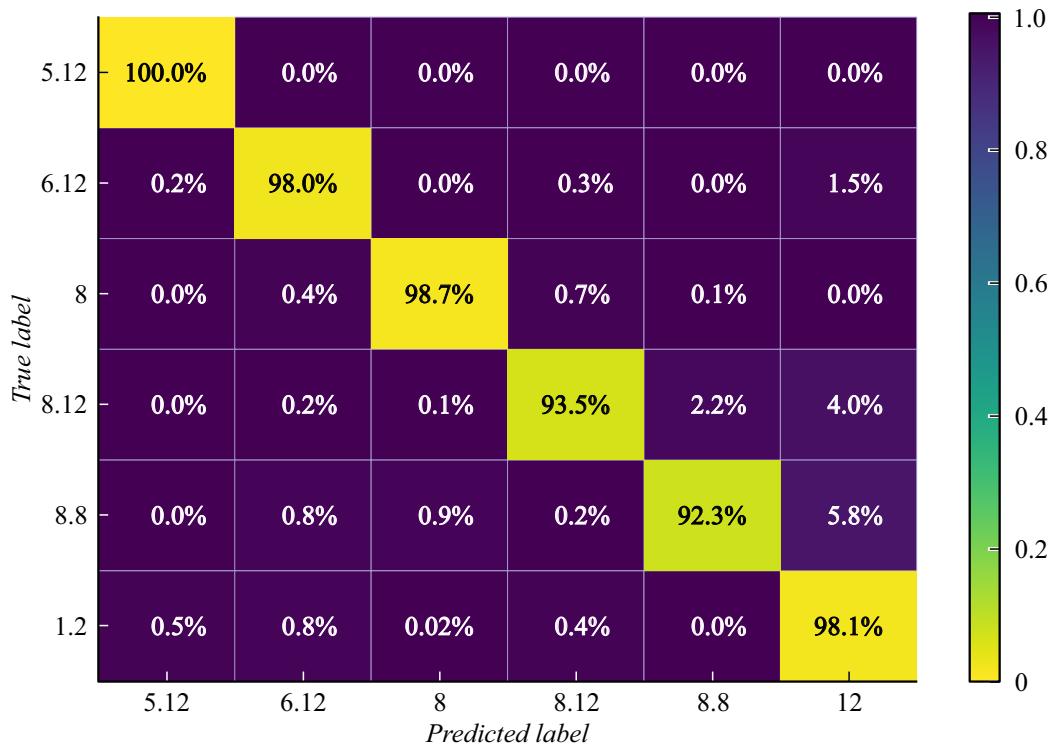


Рис. 7. Нормализованная матрица ошибок для нейронной сети, обученной на тренировочном наборе данных Cu K-XANES-спектров Cu-MOR. Строки представляют собой истинные классы, а столбцы — прогнозируемые классы. Матрица ошибок была проанализирована, чтобы выявить основные трудности классификации некоторых признаков.

типа атомного строения, в частности, конкретные кольца цеолита, в которых расположены атомы меди. Для демонстрации этой возможности было выполнено обучение нейросети с использованием теоретических спектров. Точность полученного описания характеризуется высоким значением 0.97 метрики F1. В дальнейшем построенная модель МО будет применена для анализа экспериментальных Cu K-XANES-спектров.

Финансирование работы

Работа выполнена при поддержке Российского научного фонда (№ 23-22-00438) на базе Южного федерального университета.

Конфликт интересов

Авторы заявляют, что у них нет конфликта интересов.

Список литературы

- [1] S.E. Bozbag, E.M.C. Alayon, J. Pecháček, M. Nachtegaal, M. Ranocchiari, J.A. van Bokhoven. *Catal. Sci. Technol.*, **6**, 5011 (2016). DOI: 10.1039/C6CY00041J
- [2] E.M.C. Alayon, M. Nachtegaal, A. Bodi, M. Ranocchiari, J.A. van Bokhoven. *Phys. Chem. Chem. Phys.*, **17**, 7681 (2015). DOI: 10.1039/C4CP03226H
- [3] E.M.C. Alayon, M. Nachtegaal, E. Kleymenov, J.A. van Bokhoven. *Microporous Mesoporous Mater.*, **166**, 131 (2013). DOI: 10.1016/j.micromeso.2012.04.054
- [4] B.F. Sels, L.M. Kustov (ed.). *Zeolites and Zeolite-Like Materials* (Elsevier, 2016), DOI: 10.1016/C2014-0-00257-2
- [5] J.S. Woertink, P.J. Smeets, M.H. Groothaert, M.A. Vance, B.F. Sels, R.A. Schoonheydt, E.I. Solomon. *Proc. Natl. Acad. Sci.*, **106**, 18908 (2009). DOI: 10.1073/pnas.0910461106
- [6] В.В. Прядченко, Г.Б. Сухарина, А.М. Ермакова, С.В. Базовская, Т.И. Курзина, В.А. Дурыманов, В.А. Толстопятенко, В.В. Срабионян, Л.А. Авакян, Л.А. Бугаев. *ЖТФ*, **91** (7), 1132. DOI: 10.21883/JTF.2021.07.50954.166-20
[V.V. Pryadchenko, G.B. Sukharina, A.M. Ermakova, S.V. Bazovaya, T.I. Kurzina, V.A. Durymanov, V.A. Tolstopiatenko, V.V. Srabionyan, L.A. Avakyan, L.A. Bugaev. *Tech. Phys.*, **66**, 1018 (2021). DOI: 10.1134/S1063784221070124]
- [7] В.В. Срабионян, Г.Б. Сухарина, С.Ю. Каптелинин, В.А. Дурыманов, А.М. Ермакова, Т.И. Курзина, Л.А. Авакян, Л.А. Бугаев. *ФТТ*, **62** (7), 1082 (2020). DOI: 10.21883/FTT.2020.07.49477.018 [V.V. Srabionyan, G.B. Sukharina, S.Y. Kaptelinin, V.A. Durymanov, A.M. Ermakova, T.I. Kurzina, L.A. Avakyan, L.A. Bugaev. *Phys. Solid State*, **62**, 1222 (2020). DOI: 10.1134/S1063783420070252]
- [8] V.V. Srabionyan, G.B. Sukharina, T.I. Kurzina, V.A. Durymanov, A.M. Ermakova, L.A. Avakyan, E.M.C. Alayon, M. Nachtegaal, J.A. van Bokhoven, L.A. Bugaev. *J. Phys. Chem. C*, **125**, 25867 (2021). DOI: 10.1021/acs.jpcc.1c08240

- [9] S. Grundner, M.A.C. Markovits, G. Li, M. Tromp, E.A. Pidko, E.J.M. Hensen, A. Jentys, M. Sanchez-Sanchez, J.A. Lercher. *Nat. Commun.*, **6**, 7546 (2015). DOI: 10.1038/ncomms8546
- [10] Q. Zhang, J. Yu, A. Corma. *Adv. Mater.*, **32**, 2002927 (2020). DOI: 10.1002/adma.202002927
- [11] A.A. Guda, S.A. Guda, A. Martini, A.L. Bugaev, M.A. Soldatov, A.V. Soldatov, C. Lamberti. *Radiat. Phys. Chem.*, **175**, 108430 (2020). DOI: 10.1016/j.radphyschem.2019.108430
- [12] A.A. Guda, S.A. Guda, A. Martini, A.N. Kravtsova, A. Algasov, A. Bugaev, S.P. Kubrin, L.V. Guda, P. Šot, J.A. van Bokhoven, C. Copéret, A.V. Soldatov. *Npj Comput. Mater.*, **7**, 203 (2021). DOI: 10.1038/s41524-021-00664-9
- [13] L. Avakyan, D. Tolchina, V. Barkovski, S. Belenov, A. Alekseenko, A. Shaginyan, V. Srabionyan, V. Guterman, L. Bugaev. *Comput. Mater. Sci.*, **208**, 111326 (2022). DOI: 10.1016/j.commatsci.2022.111326
- [14] M. Moliner, Y. Román-Leshkov, A. Corma. *Acc. Chem. Res.*, **52**, 2971 (2019). DOI: 10.1021/acs.accounts.9b00399
- [15] A. Gandhi, M.M.F. Hasan. *Curr. Opin. Chem. Eng.*, **35**, 100739 (2022). DOI: 10.1016/j.coche.2021.100739
- [16] J.D. Evans, F.-X. Coudert. *Chem. Mater.*, **29**, 7833 (2017). DOI: 10.1021/acs.chemmater.7b02532.
- [17] J.H. Friedman. *Ann. Stat.*, **29**, 1189 (2001). DOI: 10.1214/aos/1013203451
- [18] T. Hastie, J. Friedman, R. Tibshirani. *The Elements of Statistical Learning* (Springer, NY, 2001), DOI: 10.1007/978-0-387-21606-5
- [19] Q. Zhu, Y. Gu, X. Liang, X. Wang, J. Ma. *ACS Catal.*, **12**, 12336 (2022). DOI: 10.1021/acscatal.2c03250
- [20] A.V. Soldatov, A.N. Kravtsova, L.N. Mazalov, S.V. Trubina, N.A. Kryuchkova, G.B. Sukharina. *J. Struct. Chem.*, **48**, 1061 (2007). DOI: 10.1007/s10947-007-0171-0
- [21] J.J. Rehr, A.L. Ankudinov. *Coord. Chem. Rev.*, **249**, 131 (2005). DOI: 10.1016/j.ccr.2004.02.014
- [22] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I.E. Castelli, R. Christensen, M. Dułak, J. Friis, M.N. Groves, B. Hammer, C. Hargas, E.D. Hermes, P.C. Jennings, P. Bjerre Jensen, J. Kermode, J.R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K.S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, K.W. Jacobsen. *J. Phys. Condens. Matter*, **29**, 273002 (2017). DOI: 10.1088/1361-648X/aa680e
- [23] G.M. Psolfogiannakis, J.F. McCleerey, E. Jaramillo, A.C.T. van Duin. *J. Phys. Chem. C*, **119**, 6678 (2015). DOI: 10.1021/acs.jpcc.5b00699
- [24] Y. Joly. *Phys. Rev. B*, **63**, 125120 (2001). DOI: 10.1103/PhysRevB.63.125120
- [25] M.S. Asyaky, R. Mandala. *Improving the Performance of HDBSCAN on Short Text Clustering by Using Word Embedding and UMAP*, in: 2021 8th Int. Conf. Adv. Informatics Concepts, Theory Appl., IEEE, (2021), p. 6. DOI: 10.1109/ICAICTA53211.2021.9640285
- [26] B.E.R. Snyder, P. Vanelderen, R.A. Schoonheydt, B.F. Sels, E.I. Solomon. *J. Am. Chem. Soc.*, **140**, 9236 (2018). DOI: 10.1021/jacs.8b05320
- [27] M.H. Mahyuddin, T. Tanaka, A. Staykov, Y. Shiota, K. Yoshizawa. *Inorg. Chem.*, **57**, 10146 (2018). DOI: 10.1021/acs.inorgchem.8b01329
- [28] P. Han, Z. Zhang, Z. Chen, J. Lin, S. Wan, Y. Wang, S. Wang. *Catalysts*, **11**, 751 (2021). DOI: 10.3390/catal11060751
- [29] Электронный ресурс. Режим доступа: <https://pytorch.org/>